

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 07:24:19 ; Search time 37 Seconds
(without alignments)
1310.898 Million cell updates/sec

Title: US-09-619-032-4
Perfect score: 1880
Sequence: 1 MRALVFNHLYAEIPKSEI.....RLDAFRAIYNDRGENGEP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1874	99.7	364	18 AAW34643	Thermotable alpha
2	645	34.3	128	23 AAC21657	AEDII12RA alpha-ga
3	142.5	7.6	649	15 AAR47504	Pyrococcus furiosu
4	141.5	7.5	655	22 AAB96088	Putative P. abyssal
5	126	6.7	1362	22 AAB96091	Putative P. abyssal
6	122.5	6.5	653	19 AAW54870	Super heat resista
7	103	5.5	571	21 AAG16038	Arabidopsis thalia
8	103	5.5	571	21 AAG48655	Arabidopsis thalia
9	103	5.5	704	21 AAG16037	Arabidopsis thalia
10	103	5.5	704	21 AAG48654	Arabidopsis thalia

11	103	5.5	726	21 AAG16036	Arabidopsis thalia
12	103	5.5	726	21 AAG48653	Arabidopsis thalia
13	103	5.5	1073	22 AAU36768	Staphylococcus aur
14	103	5.5	1147	22 AAU33972	Staphylococcus aur
15	101.5	5.4	360	22 AAG82528	S. epidermidis ope
16	101.5	5.4	376	22 AAG82506	S. epidermidis ope
17	101.5	5.4	936	22 AAG18270	Novel human diagno
18	101	5.4	329	17 AAG94013	Heat resistant mal
19	100	5.3	929	22 ABB70857	Drosophila melanog
20	94.5	5.0	367	23 ABP38833	Staphylococcus epi
21	93.5	5.0	1320	23 AAM47603	Drosophila cell cy
22	93	4.9	216	21 AAG29078	Arabidopsis thalia
23	93	4.9	218	21 AAG29077	Arabidopsis thalia
24	93	4.9	235	21 AAG29076	Arabidopsis thalia
25	93	4.9	4829	22 AAG97833	Human apoptosis-in
26	92.5	4.9	392	23 ABB54675	Lactococcus lactis
27	92.5	4.9	625	22 ABB61213	Drosophila melanog
28	92	4.9	375	22 AAG82373	S. epidermidis ope
29	91.5	4.9	480	21 AAB16318	Pinus radiata UGP
30	91.5	4.9	1009	23 AAO20949	Protein of the Bpm
31	91	4.8	364	21 AAG14039	Arabidopsis thalia
32	91	4.8	384	21 ABB92380	Herbicidally activ
33	91	4.8	389	21 AAG14038	Arabidopsis thalia
34	91	4.8	1248	22 ABB61465	Drosophila melanog
35	91	4.8	1417	20 AAY31947	Plasmodium falci
36	90.5	4.8	317	22 AAB96470	Putative P. abyssal
37	90.5	4.8	406	23 ABP38725	Staphylococcus epi
38	89.5	4.8	1688	22 AAU34158	Staphylococcus aur
39	89.5	4.8	2397	22 AAB36672	Staphylococcus aur
40	89.5	4.8	2766	22 AAB83972	Amino acid sequen
41	89	4.7	292	23 ABB54281	Lactococcus lactis
42	89	4.7	1039	22 AAG15145	Novel human diagno
43	89	4.7	1073	22 AAG90611	C glutamicum prote
44	88.5	4.7	473	21 AAG07943	A uridine diphosph
45	88.5	4.7	473	21 AAG34044	Zea mays protein f

ALIGNMENTS

RESULT 1
AAW34643
ID AAW34643 standard; Protein: 364 AA.
XX
AC AAW34643;
XX
DT 27-MAR-1998 (first entry)
XX
DE Thermotable alpha-galactosidase AEDII12RA-alpha-gal-186C.
XX
KW Alpha-galactosidase; alpha-galactosidase; thermostable enzyme;
KW food processing; alpha glycoside hydrolysis; raffinose;
KW stachyose; verbascose; bean; flatulence; AEDII12RA-alpha-gal-186C.
XX
OS Thermococcus alcaliphilus strain AEDII12RA.
XX
FH Key Location/Qualifiers
FT Misc-difference 329
ET /note= "encoded by CTT"
XX
PN WO9732974-A1.
XX
PD 12-SEP-1997.
XX
PF 05-FEB-1997; 97MO-US01452.
XX
PR 08-MAR-1996; 96US-0613220.
XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Murphy D, Reid J, Rudolph MJ;
XX
DR WPI; 1997-470541/43.

DR N-PSDB; AAT93753.
 XX
 PT Nucleic acid encoding alpha-galactosidase from Thermococcus
 PT alcaliphilus - used in food processing to hydrolyse
 PT alpha-glycosides, e.g. raffinose
 XX
 PS Claim 1; Fig 1; 32pp; English.
 XX
 CC This protein comprises AEDII12RA-alpha-gal-18GC, a claimed
 CC thermostable alpha-galactosidase of Thermococcus alcaliphilus
 CC AEDII12RA, a bacterium that shows optimum growth at 85 deg C and
 CC pH 9.5. Also claimed are: (1) an isolated polynucleotide (see
 CC AAT93753) encoding the alpha-galactosidase; (2) a vector containing
 CC the polynucleotide or homologous or complementary sequences; (2)
 CC host cells containing the vector; (3) a process for producing the
 CC alpha-galactosidase in transformed or transfected host cells; an
 CC enzyme showing at least 70% identity to alpha-galactosidase and
 CC comprising at least 30 amino acid residues of its sequence; and (4)
 CC a method for hydrolysing alpha-galactoside bonds using the enzyme.
 CC The enzyme can be used to hydrolyse raffinose to sucrose and glucose
 CC in sugar beet processing (raffinose inhibits crystallisation of
 CC sucrose), and as a digestive aid to hydrolyse raffinose, stachyose
 CC and verbascose in beans and other gassy foods.
 XX
 SQ Sequence 364 AA;
 Query Match 99.7%; Score 1874; DB 18; Length 364;
 Best Local Similarity 99.5%; Pred. No. 2.3e-180;
 Matches 362; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRALVFHGNLOYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
 Db :|||||
 Db 1 LRALVFHGNLOYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
 QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPELAY 120
 Db :|||||
 Db 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPELAY 120
 QY 121 DPIIPAILKNDNGYEYLFADGEAMLFSAHLSAIIKPIKLYPHLIKAQREKFRYSYLLG 180
 Db :|||||
 Db 121 DPIIPAILKNDNGYEYLFADGEAMLFSAHLSAIIKPIKLYPHLIKAQREKFRYSYLLG 180
 QY 181 LRELKAIKLVPEGKVTLKAVKDIEAPVWVAVNTAVMLGIGRLPLMPKPKVASHIEDKD 240
 Db :|||||
 Db 181 LRELKAIKLVPEGKVTLKAVKDIEAPVWVAVNTAVMLGIGRLPLMPKPKVASHIEDKD 240
 QY 241 NILLYGTDTIEFTGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
 Db :|||||
 Db 241 NILLYGTDTIEFTGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
 QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDRGE 360
 Db :|||||
 Db 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDRGE 360
 QY 361 NGEF 364
 Db :|||
 Db 361 NGEF 364
 RESULT 2
 ID AAO21657
 XX AAO21657 standard; Protein; 128 AA.
 AC AAO21657;
 XX
 DT 19-SEP-2002 (first entry)
 DE AEDII12RA alpha-galactosidase 18GC protein.
 XX
 KW Alpha-galactosidase; chromosomal walking; hydrolysis; saccharide;
 KW catalyst; Thermococcus alcaliphilus; AEDII12RA alpha-galactosidase 18GC;
 KW enzyme.
 XX

OS Thermococcus alcaliphilus.
 XX
 PN US2002045226-A1.
 XX
 PD 18-APR-2002.
 XX
 PF 20-JUN-2001; 2001US-0886400.
 XX
 PR 08-MAR-1996; 96US-0613220.
 PR 28-SEP-1999; 99US-0407806.
 PR 19-JUL-2000; 2000US-0619032.
 XX
 PA (MURP/) MURPHY D.
 PA (REID/) REID J.
 XX
 PI Murphy D, Reid J;
 XX
 DR WPI; 2002-507240/54.
 DR N-PSDB; AAL39618.
 XX
 PT New nucleic acid encoding heat-stable alpha-galactosidase, and related
 PT polypeptides and antibodies, useful for hydrolysis of saccharides -
 XX
 PS Disclosure; Fig 5; 35pp; English.
 XX
 CC The invention relates to an isolated nucleic acid that comprises a
 CC sequence not reproduced, variants with at least 50% identity, and encoded
 CC polypeptides with alpha-galactosidase activity. The polynucleotides of
 CC the invention can be used as a source of probes and primers for
 CC chromosomal walking and for identifying related sequences. The protein of
 CC the invention is used to catalyse the hydrolysis of saccharides; to
 CC modify libraries of small molecules which are then screened for
 CC particular activities and to generate specific antibodies, useful for
 CC detection or purification of the protein of the invention and related
 CC peptides. This sequence represents the Thermococcus alcaliphilus
 CC AEDII12RA alpha-galactosidase 18GC of the invention.
 XX
 SQ Sequence 128 AA;
 Query Match 34.3%; Score 645; DB 23; Length 128;
 Best Local Similarity 99.2%; Pred. No. 5.5e-57;
 Matches 127; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRALVFHGNLOYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
 Db :|||||
 Db 1 LRALVFHGNLOYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
 QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPELAY 120
 Db :|||||
 Db 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPELAY 120
 QY 121 DPIIPAIL 128
 Db :|||||
 Db 121 DPIIPAIL 128
 RESULT 3
 ID AAR47504
 XX AAR47504 standard; protein; 649 AA.
 AC AAR47504;
 XX
 DT 07-JUL-1994 (first entry)
 DE Pyrococcus furiosus alpha amylase.
 XX
 KW Pyrococcus furiosus; alpha amylase; liquefaction; polymers;
 KW glucopolymers; thermostable.
 XX
 OS Pyrococcus furiosus.
 XX
 PN EP577257-A.
 XX

RESULT 7
 AAG16038
 ID AAG16038 standard; Protein; 571 AA.
 XX
 AC AAG16038;
 DT
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 16524.
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 OS
 PN EP1033405-A2.
 PD
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 14-JUN-1999; 99US-0138847.
 PR 16-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140981.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 19-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 21-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.

```
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 5.5%; Score 103; DB 21; Length 571;
Matches 84; Conservative 45; Mismatches 110; Indels 130; Gaps 20;

QY 29 IPVITLKEEIPFGLNTGTTLKPLPKDIIDLKGGTASDL-----IEITGTSYTHAIL 83
Db 101 LPILF-MVPEELHF---TPVYALSFIMGDPDKMATLGDINOLPTGVKIEKLQRLLTKTML 156
QY 84 PLL-----PLSRVQAQVQRDEVEKEELFEVSPKGFWLP 118
Db 157 PLLSELGIIIPREYLLWKLLRSGCAYANSRIHA-VQ-----AEVLVLASGKMDMLPSQ 210
QY 119 AVDPIIPAIL-----KNGCYELFADGFAMLFSAHLNSAIKPIKPLYPHLIKAOREK 170
```

```
Db 211 BEAKRLHGLLKNCVRCFKDNGHTLLLEDLSILL-----TVIKCTG-----KYRSW 257
QY 171 PERYIS-----YLLG--LRELKAKILVP-----EGKVTLKAVKDIEAVP---V 209
Db 258 RYDLVSDFLPFSKGLAYALDEVLGFLRNAVGSVFFSTMEDGKI-----VKGLAGVPDKGP 313
QY 210 WVAVNTAVMLGIGRLPLMNPKKVASWIEDKNILL-----YGTIDIE 250
Db 314 VLLVGYHMLGLELGPMS-----AFIKER-NILFRGMAHPVLYSDNDPAKAFDYGWIK 367
QY 251 FIGYRDIAGYRMSVEGLLEVIDELNSELCLPS---ELKHSGRELYLRTSSWAPDKSLRIW 307
Db 368 VFG-----AYPVTANLFLKLLDSKSHVLLPFGGAREALHNRGEY-----KLIW 411
QY 308 REDEGNARL 316
Db 412 PEQQEFVRM 420

RESULT 8
AAG48655
ID AAG48655 standard; Protein; 571 AA.
XX AAG48655;
AC AAG48655;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61465.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135253.
PR 24-MAY-1999; 99US-0135629.
```

us-09-619-032-4.rag

Mon Nov '18 08:39:33 2002

PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139899.	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	23-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	99US-0142803.	PR	10-SEP-1999;	99US-0153070.
PR	28-JUN-1999;	99US-0140823.	PR	13-SEP-1999;	99US-0153758.
PR	29-JUN-1999;	99US-0140991.	PR	15-SEP-1999;	99US-0154018.
PR	30-JUN-1999;	99US-0141287.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0141842.	PR	20-SEP-1999;	99US-0154779.
PR	01-JUL-1999;	99US-0142154.	PR	22-SEP-1999;	99US-0155139.
PR	02-JUL-1999;	99US-0142055.	PR	23-SEP-1999;	99US-0155486.
PR	06-JUL-1999;	99US-0142390.	PR	24-SEP-1999;	99US-0155659.
PR	08-JUL-1999;	99US-0142803.	PR	28-SEP-1999;	99US-0156458.
PR	09-JUL-1999;	99US-0142920.	PR	29-SEP-1999;	99US-0156596.
PR	12-JUL-1999;	99US-0142977.	PR	04-OCT-1999;	99US-0157117.
PR	13-JUL-1999;	99US-0143547.	PR	05-OCT-1999;	99US-0157753.
PR	14-JUL-1999;	99US-0143624.	PR	06-OCT-1999;	99US-0157865.
PR	15-JUL-1999;	99US-0144005.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144088.	PR	08-OCT-1999;	99US-0158232.
PR	16-JUL-1999;	99US-0144086.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159330.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159638.
PR	20-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160768.
PR	21-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145192.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145218.	PR	25-OCT-1999;	99US-0160989.
PR	23-JUL-1999;	99US-0145224.	PR	25-OCT-1999;	99US-0161404.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145918.	PR	26-OCT-1999;	99US-0161359.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161360.
PR	28-JUL-1999;	99US-0145951.	PR	28-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146386.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161992.
PR	02-AUG-1999;	99US-0146389.			

```

PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      5.5%  Score 103; DB 21; Length 571;
Best Local Similarity 22.8%; Pred. No. 0.29;
Matches 84; Conservative 45; Mismatches 110; Indels 130; Gaps 20;

QY 29 IPVETLIKIEIPGLNITGYTKFLPKDIDLVKGIASDL-----IEIIGTSYTHAL 83
Db 101 LPILF-MVPEELHF---TVPYALSFIMGDDPKMATLGIDNQLPTGVKIEKLQRLLTKML 156
QY 84 PLL-----PLSRVFAQVORDREVKELEFVSPKGFWLPEL 118
Db 157 PLLSELGIIIPRETLMLKLLKSGCAYANSRIHA-VQ-----AEVILVILASGKDMMLPSQ 210
QY 119 AYDPIIPAIL-----KONGYEYLFADGEAMLFSAHLNSAIKPIKPLYPLHKAQREK 170
Db 211 EEAKRLHGLLKNCVRCFKDNGHTLLLEDISILL-----TVIKGTG-----KYRSW 257
QY 171 RFRVIS-----YLLG--LRELRAIKLVF-----EGKVTLKAVKDIEAVP---V 209
Db 258 RYDLVDSFLPPSGELAYALDEVLGFLRNAVGSVFSTMEDGKI---VKGLAGVDPKGP 313
QY 210 WVAVNTAVMLGIGRLPLMNPCKVASWIEDKDNILL-----YGTDIE 250
Db 314 VLLVGYHMLGLGLGPSE-----AFIKK-NILFRGMHPVLYSDNDPAKAFDYGDWIK 367
QY 251 FIGYRDIAGYRMSVEGLLEVIDELNSELCLPS---ELKHSRGRELYLRTSSWAPDKSLRIW 307
Db 368 VFG-----AYPVTATNLKILDSKSHVLLFPGAREALHNRGEQY-----KLIW 411
QY 308 REDEGNARL 316
Db 412 PEQOEVRM 420

RESULT 9
AAG16037
ID AAG16037 standard; Protein; 704 AA.
AC AAG16037;
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 16523.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
PN EP1033405-A2.
PD 06-SEP-2000.
PE 25-FEB-2000; 2000EP-0301439.
PR 05-MAR-1999; 99US-0121825.
PR 09-MAR-1999; 99US-0123180.
PR 23-MAR-1999; 99US-0123548.
PR 29-MAR-1999; 99US-0125788.
PR 01-APR-1999; 99US-0126264.
PR 06-APR-1999; 99US-0126785.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 28-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.

```


PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	22-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 5.5%; Score 103; DB 21; Length 704;		
Best Local Similarity 22.8%; Pred. No. 0.4; Indels 130; Gaps 20;		
Matches 84; Conservative 45; Mismatches 110; Indels 130; Gaps 20;		
Qy	29	IPVETLIKKEIPFGLNITGYTLKFLPKDIIDLVKGGIASDL-----IEIGTSYTHAIL 83
Db	234	LPILF-MVPEELHF--TVPYALSFINGDPKMATLIGDNLQPTGVKIEKLRQLTKTML 289
Qy	84	PLL-----PLSRVEAQVORDREYKEELFEVSPKGFWLPPEL 118
Db	290	PLLSELGGIIPRETLKWLKLLRSGCAYANSRIHA-VO-----AEVLVLASGKDMMLPSQ 343
Qy	119	AYDPIIPAIL-----KNGVEYLFADGEAMLFSAHLSAIPKIPKPLYPHLIKAQREK 170
Db	344	EEAKRLHGLLKNCVRCFKNGHWHLLLEDLSISLL-----TVIKGTG-----KYRRSW 390
Qy	171	RFYRIS-----YLLG--LRELKAIKLVF-----EGKVTLKAVKDIEAVP---V 209
Db	391	RYDLVSDFLPSPKGEALAYALDEVILGFLRNAVGSVFFSTMEDGKI---VRGLAGVPDKGP 446
Qy	210	WVAVNTAVMLGIGRPLPMNPKKVASWIEDKNILL-----YGTDIE 250
Db	447	VLLGVYHMLGLEIGPMSE-----AFIKER-NILFRGMHPVLYSDNDPAKAFDYGDWIK 500
Qy	251	FIGYRDIAGYRMSVEGLEVIDELNSLCLPS---ELKHSGRELYLRTSSWAPDKSLRIW 307
Db	501	VFG-----AYPTATNLFKLLDSKSHVLLFPGGAREALHNRGEY-----KLIW 544
Qy	308	REDEGNARL 316
Db	545	PEQGEFVRM 553
RESULT 10		
AAG48654		
ID	AAG48654	standard; Protein; 704 AA.
XX	XX	
AC	AAG48654;	
XX	XX	
DT	18-OCT-2000	(first entry)
XX	XX	
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 61464.
XX	XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
XX	termination sequence.	
OS	Arabidopsis thaliana.	
XX	XX	
PN	EP1033405-A2.	
XX	XX	
PD	06-SEP-2000.	
XX	XX	

PF XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 28-APR-1999; 99US-0130510.
PR 30-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132486.
PR 14-MAY-1999; 99US-0132556.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149920.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

PR	18-JUN-1999;	99US-0139462
PR	18-JUN-1999;	99US-0139463
PR	18-JUN-1999;	99US-0139464
PR	18-JUN-1999;	99US-0139750
PR	21-JUN-1999;	99US-0139763
PR	21-JUN-1999;	99US-0139817
PR	22-JUN-1999;	99US-0139899
PR	23-JUN-1999;	99US-0140353
PR	23-JUN-1999;	99US-0140354
PR	24-JUN-1999;	99US-0140695
PR	28-JUN-1999;	99US-0140823
PR	29-JUN-1999;	99US-0140991
PR	30-JUN-1999;	99US-0141287
PR	01-JUL-1999;	99US-0141842
PR	01-JUL-1999;	99US-0142154
PR	02-JUL-1999;	99US-0142055
PR	06-JUL-1999;	99US-0142390
PR	08-JUL-1999;	99US-0142803
PR	09-JUL-1999;	99US-0142920
PR	12-JUL-1999;	99US-0142977
PR	13-JUL-1999;	99US-0143542
PR	14-JUL-1999;	99US-0143624
PR	15-JUL-1999;	99US-0144005
PR	16-JUL-1999;	99US-0144085
PR	16-JUL-1999;	99US-0144086
PR	19-JUL-1999;	99US-0144325
PR	19-JUL-1999;	99US-0144331
PR	19-JUL-1999;	99US-0144332
PR	19-JUL-1999;	99US-0144333
PR	19-JUL-1999;	99US-0144334
PR	20-JUL-1999;	99US-0144335
PR	20-JUL-1999;	99US-0144352
PR	20-JUL-1999;	99US-0144632
PR	20-JUL-1999;	99US-0144684
PR	21-JUL-1999;	99US-0144814
PR	21-JUL-1999;	99US-0145086
PR	21-JUL-1999;	99US-0145088
PR	22-JUL-1999;	99US-0145085
PR	22-JUL-1999;	99US-0145087
PR	22-JUL-1999;	99US-0145089
PR	22-JUL-1999;	99US-0145192
PR	23-JUL-1999;	99US-0145145
PR	23-JUL-1999;	99US-0145218
PR	23-JUL-1999;	99US-0145224
PR	26-JUL-1999;	99US-0145724
PR	27-JUL-1999;	99US-0145913
PR	27-JUL-1999;	99US-0145916
PR	27-JUL-1999;	99US-0145918
PR	28-AUG-1999;	99US-0145951
PR	02-AUG-1999;	99US-0146386
PR	02-AUG-1999;	99US-0146388
PR	02-AUG-1999;	99US-0146389
PR	03-AUG-1999;	99US-0147038
PR	04-AUG-1999;	99US-0147202
PR	04-AUG-1999;	99US-0147304
PR	05-AUG-1999;	99US-0147192
PR	05-AUG-1999;	99US-0147260
PR	06-AUG-1999;	99US-0147303
PR	06-AUG-1999;	99US-0147416
PR	08-AUG-1999;	99US-0147493
PR	09-AUG-1999;	99US-0147935
PR	10-AUG-1999;	99US-0148171
PR	11-AUG-1999;	99US-0148319
PR	12-AUG-1999;	99US-0148341
PR	13-AUG-1999;	99US-0148561
PR	13-AUG-1999;	99US-0148565
PR	16-AUG-1999;	99US-0148684
PR	17-AUG-1999;	99US-0149368
PR	18-AUG-1999;	99US-0149175
PR	18-AUG-1999;	99US-0149426
PR	20-AUG-1999;	99US-0149722
PR	20-AUG-1999;	99US-0149723
PR	20-AUG-1999;	99US-0149929
PR	23-AUG-1999;	99US-0149902
PR	23-AUG-1999;	99US-0149930

PR	25-AUG-1999;	99US-0150566
PR	26-AUG-1999;	99US-0150584
PR	27-AUG-1999;	99US-0151065
PR	27-AUG-1999;	99US-0151066
PR	27-AUG-1999;	99US-0151080
PR	30-AUG-1999;	99US-0151303
PR	31-AUG-1999;	99US-0151438
PR	01-SEP-1999;	99US-0151930
PR	07-SEP-1999;	99US-0152363
PR	10-SEP-1999;	99US-0153070
PR	13-SEP-1999;	99US-0153758
PR	15-SEP-1999;	99US-0154018
PR	16-SEP-1999;	99US-0154033
PR	20-SEP-1999;	99US-0154779
PR	22-SEP-1999;	99US-0155139
PR	23-SEP-1999;	99US-0155486
PR	24-SEP-1999;	99US-0155659
PR	28-SEP-1999;	99US-0156458
PR	04-OCT-1999;	99US-0156596
PR	09-OCT-1999;	99US-0157117
PR	05-OCT-1999;	99US-0157753
PR	06-OCT-1999;	99US-0157865
PR	07-OCT-1999;	99US-0158029
PR	08-OCT-1999;	99US-0158232
PR	12-OCT-1999;	99US-0158369
PR	13-OCT-1999;	99US-0159293
PR	13-OCT-1999;	99US-0159294
PR	14-OCT-1999;	99US-0159295
PR	14-OCT-1999;	99US-0159329
PR	14-OCT-1999;	99US-0159330
PR	14-OCT-1999;	99US-0159331
PR	14-OCT-1999;	99US-0159637
PR	14-OCT-1999;	99US-0159638
PR	18-OCT-1999;	99US-0159584
PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160814
PR	21-OCT-1999;	99US-0160815
PR	22-OCT-1999;	99US-0160980
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0160989
PR	25-OCT-1999;	99US-0161404
PR	25-OCT-1999;	99US-0161405
PR	25-OCT-1999;	99US-0161406
PR	26-OCT-1999;	99US-0161359
PR	26-OCT-1999;	99US-0161360
PR	26-OCT-1999;	99US-0161361
PR	28-OCT-1999;	99US-0161920
PR	28-OCT-1999;	99US-0161992
PR	28-OCT-1999;	99US-0161993
PR	29-OCT-1999;	99US-0162142

Query Match 5.5%; Score 103; DB 21; Length 726;

Best Local Similarity	22.8%; Pred. NO. 0.42;	
Matches	84; Conservative	45; Mismatches 110; Indels 130; Gaps 20;

[illegible]

CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1073 AA;

Query Match 5.5%; Score 103; DB 22; Length 1073;
Best Local Similarity 21.2%; Pred. No. 0.77; Indels 78; Gaps 15;
Matches 58; Conservative 48; Mismatches 90;

QY 107 EVSPKGFWLPYLAIPAILKONGYEYLFADGEAMLFSAHLNSAIKPIKPLYPHLIKA 166
Db 107 EVSPKGFWLPYLAIPAILKONGYEYLFADGEAMLFSAHLNSAIKPIKPLYPHLIKA 166
QY 309 EISP-----YYDSL---LVKLSTHAISFKQAEKMWRSRLRMIRGVKTNIPFLINV 357
Db 309 EISP-----YYDSL---LVKLSTHAISFKQAEKMWRSRLRMIRGVKTNIPFLINV 357
QY 167 QREKRF-----RYISYLLGLRELKAI-----KLVPFGKVTLLKAVKDIE----- 205
Db 167 QREKRF-----RYISYLLGLRELKAI-----KLVPFGKVTLLKAVKDIE----- 205
QY 358 MNKKFTSGDYTKFTTEETPELFDIOPSLDRGKTLETVIGNVTINGFPNVEKRPKPDYEL 417
Db 358 MNKKFTSGDYTKFTTEETPELFDIOPSLDRGKTLETVIGNVTINGFPNVEKRPKPDYEL 417
QY 206 -AVPVWVAVNTAVMLGIGR-LPLMNPKKVASWIEDKDNILLYGTDIEF-IGYRDIAGYRM 262
Db 206 -AVPVWVAVNTAVMLGIGR-LPLMNPKKVASWIEDKDNILLYGTDIEF-IGYRDIAGYRM 262
QY 418 ASIPTVSSSKIASFSFGTKQLLDEVGPKGVAEWKKQDDVLL--TDTTFRAHQSLLATRV 475
Db 418 ASIPTVSSSKIASFSFGTKQLLDEVGPKGVAEWKKQDDVLL--TDTTFRAHQSLLATRV 475
QY 263 SVGELLEVDELNSLCLPSELKHSRGRELYLRTSSWAPDK-SLRIRWDEGNARLNMLSY 321
Db 263 SVGELLEVDELNSLCLPSELKHSRGRELYLRTSSWAPDK-SLRIRWDEGNARLNMLSY 321
QY 476 RTKDMINIAS-----KTADVFKDGFSLMW-----GGATFD-VAY 509
Db 476 RTKDMINIAS-----KTADVFKDGFSLMW-----GGATFD-VAY 509
QY 322 NMRGELAFLAENSDARGWEPLPERRLDAPRAIYN 355
Db 322 NMRGELAFLAENSDARGWEPLPERRLDAPRAIYN 355
QY 510 N-----FLKENP-----WERLERL-----KAIPN 529
Db 510 N-----FLKENP-----WERLERL-----KAIPN 529

RESULT 14
AAU33972
ID AAU33972 standard; Protein; 1147 AA.
XX
AC AAU33972;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #248.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN W0200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS51831.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 5468; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence represents an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1147 AA;

Query Match 5.5%; Score 103; DB 22; Length 1147;
Best Local Similarity 21.2%; Pred. No. 0.85;
Matches 58; Conservative 48; Mismatches 90; Indels 78; Gaps 15;

QY 107 EVSPKGFWLPYLAIPAILKONGYEYLFADGEAMLFSAHLNSAIKPIKPLYPHLIKA 166
Db 107 EVSPKGFWLPYLAIPAILKONGYEYLFADGEAMLFSAHLNSAIKPIKPLYPHLIKA 166
QY 385 EISP-----YYDSL---LVKLSTHAISFKQAEKMWRSRLRMIRGVKTNIPFLINV 433
Db 385 EISP-----YYDSL---LVKLSTHAISFKQAEKMWRSRLRMIRGVKTNIPFLINV 433
QY 167 QREKRF-----RYISYLLGLRELKAI-----KLVPFGKVTLLKAVKDIE----- 205
Db 167 QREKRF-----RYISYLLGLRELKAI-----KLVPFGKVTLLKAVKDIE----- 205
QY 434 MNKKFTSGDYTKFTTEETPELFDIOPSLDRGKTLETVIGNVTINGFPNVEKRPKPDYEL 493
Db 434 MNKKFTSGDYTKFTTEETPELFDIOPSLDRGKTLETVIGNVTINGFPNVEKRPKPDYEL 493
QY 206 -AVPVWVAVNTAVMLGIGR-LPLMNPKKVASWIEDKDNILLYGTDIEF-IGYRDIAGYRM 262
Db 206 -AVPVWVAVNTAVMLGIGR-LPLMNPKKVASWIEDKDNILLYGTDIEF-IGYRDIAGYRM 262
QY 494 ASIPTVSSSKIASFSFGTKQLLDEVGPKGVAEWKKQDDVLL--TDTTFRAHQSLLATRV 551
Db 494 ASIPTVSSSKIASFSFGTKQLLDEVGPKGVAEWKKQDDVLL--TDTTFRAHQSLLATRV 551
QY 263 SVGELLEVDELNSLCLPSELKHSRGRELYLRTSSWAPDK-SLRIRWDEGNARLNMLSY 321
Db 263 SVGELLEVDELNSLCLPSELKHSRGRELYLRTSSWAPDK-SLRIRWDEGNARLNMLSY 321
QY 552 RTKDMINIAS-----KTADVFKDGFSLMW-----GGATFD-VAY 585
Db 552 RTKDMINIAS-----KTADVFKDGFSLMW-----GGATFD-VAY 585
QY 322 NMRGELAFLAENSDARGWEPLPERRLDAPRAIYN 355
Db 322 NMRGELAFLAENSDARGWEPLPERRLDAPRAIYN 355
QY 586 N-----FLKENP-----WERLERL-----KAIPN 605
Db 586 N-----FLKENP-----WERLERL-----KAIPN 605

RESULT 15
AAG82528
ID AAG82528 standard; Protein; 360 AA.
XX
AC AAG82528;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2150.
XX

Search completed: November 18, 2002, 07:25:09
Job time : 41 secs

KW Staphylococcus epidermidis SK1 strain; infection; diagnosis;
XX vaccination; endocarditis.

OS Staphylococcus epidermidis.
XX WO200134809-A2.
XX
XX
PD 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX (GLAX) GLAXO GROUP LTD.

XX Kimmerly WJ;

XX WPI; 2001-316495/33.

XX N-PSDB; AAH53378.

XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -

XX Claim 18; Page 578-579; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 360 AA;

Query Match 5.4%; Score 101.5; DB 22; Length 360;
Best Local Similarity 19.1%; Pred. No. 0.2;
Matches 57; Conservative 52; Mismatches 117; Indels 73; Gaps 12;

QY 4 LVFHGNIQYAEIPKSEIPKVEIKAYIPVIEFWLKEETPFGLNITGYTLKFLPKDIDLVK 63

Db 61 LIFKG-----VKRIVEDGYGSIKRLQNNI--NLIALHTNLDVNPKNRMLA 107

QY 64 GGIASDLIEIIGT--SYTHAILPLPLSRVEAQVQRDREVKKEELFEVSPKFWLPCLAYD 121

Db 108 DOIGLEINISMINTSYKYKQTFIPKNYIE-----DFKDSLNEL----- 147

QY 122 PIIPAILKNGYEYLFADGEAMLSAHLNSAIKPKPLYPHLIKAOEKRFYRYSYLLGL 181

Db 148 ----GLAKEGNYECFFESG-----KGQFKVGDASPYIGKLDL-----IEYV--- 187

QY 182 RELRKAIKLVF-----EGKVTLKAVKDIE--AVPV--WVAVNTAVMLGIGRLPLMNPVKV 232

Db 188 ----DEIKLEFWIKDNELEIKRAILNDHNYETPFDFIKMKESEYGLGIQNLQNTMT 243

QY 233 ASWIEDKNILLYGTDIEFYGRDIAGYRMSVGLLEV-----IDELNSELCLPSELKH 286

Db 244 LDEFSEYAKQLNIPSVRYTGOHDSPIKKVAIIGSGIGFEYKASQLGADVFTGDIKH 302

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 07:24:24 ; Search time 18 Seconds
(without alignments)
594.997 Million cell updates/sec

Title: US-09-619-032-4
Perfect score: 1880
Sequence: 1 MRALVFNHGNLQYAEIPKSEI.....RRLDAFRAIYNDWRGENGEP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pap: *
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pap: *
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pap: *
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pap: *
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pap: *
6: /cgn2_6/ptodata/1/1aa/backfiles1.pap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1587	84.4	346	2	US-08-613-220B-4
2	140.5	7.5	647	1	US-07-894-212A-8
3	140.5	7.5	649	1	US-07-894-212A-2
4	140.5	7.5	650	1	US-07-893-928A-1
5	101	5.4	329	1	US-08-270-013B-2
6	101	5.4	329	1	US-08-838-418-2
7	94.5	5.0	367	4	US-09-134-001C-3678
8	91.5	4.9	1009	4	US-09-693-146-4
9	90.5	4.8	406	4	US-09-134-001C-3570
10	88.5	4.7	764	4	US-09-235-451-36
11	88.5	4.7	2089	1	US-08-418-893D-23
12	88.5	4.7	2089	1	US-08-418-893D-24
13	88	4.7	1154	4	US-09-134-001C-3428
14	87.5	4.7	689	1	US-08-248-021A-2
15	87	4.6	235	2	US-09-141-135-2
16	87	4.6	657	4	US-09-355-166-1
17	85.5	4.5	511	4	US-09-134-078-21
18	84	4.5	636	4	US-09-564-803-237
19	83.5	4.4	891	1	US-09-134-001C-4913
20	83	4.4	1110	1	US-08-118-441-29
21	83	4.4	1110	3	US-08-338-579A-29
22	83	4.4	1110	5	PCT-US94-09851-29
23	82.5	4.4	549	2	US-08-676-279-59
24	82.5	4.4	788	2	US-08-907-166-6
25	82	4.4	303	3	US-09-002-298-1
26	81.5	4.3	806	1	US-08-451-715A-6
27	81	4.3	550	1	US-08-484-493-2

28	81	4.3	550	1	US-08-484-494-2	Sequence 2, Appli
29	81	4.3	550	2	US-08-345-212-2	Sequence 2, Appli
30	81	4.3	550	4	US-09-249-003-2	Sequence 2, Appli
31	80.5	4.3	491	1	US-08-489-733-3	Sequence 3, Appli
32	80.5	4.3	491	2	US-08-993-581B-3	Sequence 3, Appli
33	80.5	4.3	641	1	US-07-718-535-3	Sequence 3, Appli
34	80.5	4.3	641	1	US-08-161-999-3	Sequence 3, Appli
35	80.5	4.3	1676	4	US-08-487-283A-2	Sequence 9, Appli
36	80	4.3	329	4	US-09-180-827-9	Sequence 5, Appli
37	80	4.3	506	2	US-08-849-480A-5	Sequence 11, Appli
38	80	4.3	541	4	US-09-627-376-11	Sequence 2, Appli
39	80	4.3	1065	4	US-09-412-545-2	Sequence 9, Appli
40	79.5	4.2	906	2	US-08-609-230A-9	Sequence 44, Appli
41	79.5	4.2	1751	4	US-09-136-574A-44	Sequence 4606, Ap
42	79	4.2	246	4	US-09-104-068-4	Sequence 4, Appli
43	79	4.2	623	4	US-09-104-068-2	Sequence 2, Appli
44	79	4.2	637	4	US-09-104-068-2	Sequence 2, Appli
45	79	4.2	889	4	US-09-132-316-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-613-220B-4
; Sequence 4, Application US/08613220B
; Patent No. 5958751
; GENERAL INFORMATION:
; APPLICANT: Murphy, Dennis
; APPLICANT: Reid, John
; TITLE OF INVENTION: ALPHA-GALACTOSIDASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,220B
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 36,347
; REFERENCE/DOCKET NUMBER: 09010/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-68-5099
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 346 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-613-220B-4

Query Match 84.4%; Score 1587; DB 2; Length 346;
Best Local Similarity 94.8%; Pred No. 2.3e-160;
Matches 345; Conservative 1; Mismatches 0; Indels 18; Gaps 18;
QY 1 MRALVFNHGNLQYAEIPKSEIPIVETLIKKEIPFGLNITGYTLFPLPKDIID 60

Db 1 LRLVPHGNLQYAEIPKSE-PKVIKAYIPVIETLIKER-PFGLNITGYTLKFLPKDII- 57
QY 61 LVKGGIASDLIEIGTSYTHAILPLPLSRVAQVORDREVKEELFEVSPKGFWLPDELAY 120
Db 58 LVKGGIASDLIEIGTSYTHAILPLPLSRVAQVORDREVKEELFEVSPKGFWLPDELAY 114
QY 121 DPIIPAILKDNKGYEYLFADGEAMLSAHLNSAIKPIKPLPHLIKAQREKRPYISYLLG 180
Db 115 DPIIPAILKDNKGYEYLFAD- EAMLSAHLNSAIKPIKPL- PHLIKAQREKRPYISYLL- 171
QY 181 LRELRAIKLVPEGKVTLLKAVKDIEAVPVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
Db 172 LRELRAIKLVPEGKVTLLK-VKDIEAVPVAVNTAVML- IGRPLMNPKKVASWIEDK- 228
QY 241 NILLYGTDTIEFTGYRDIAGYRMSVEGLLEVIDELNSCLPSELKHSGLRELYLRTSSNAP 300
Db 229 NILLYGTDTIEFTGYRDIAG- RMSVEGLLEVIDELNSCLP- PSELKHSGLRELYLRTSSNA- 285
QY 301 DKSLRTWREDEGNARLNMNLSYNNMRLGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
Db 286 DKSLRTWREDEGNARLNMNLS- YNNMRLGELAFLAENSDARGW- PLPERRLDAFRAIYNDWRG- 342
QY 361 NCEP 364
Db 343 NCEP 346

RESULT 2

US-07-894-212A-8
; Sequence 8, Application US/07894212A
; Patent No. 5366883
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UMORI, TAKASHI
; APPLICANT: MUKAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARB & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/07/894,212A
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-894-212A-8

Query Match 7.5% Score 140.5; DB 1; Length 647;
Best Local Similarity 22.8%; Pred. No. 4.4e-06;
Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;
QY 25 EKAYIPVIETLIKKEIP---FGLNITGYTLKFLP---KDIIDLVKGGIASDLIEIGTSY 78
Db 27 EKCVWPFLETL--EYPMKVAIHTSGPLIEWLQDNRPVIDLLRSVLKRGQVEIVVAGF 84
QY 79 THAILPLPLSRVAQVORDREVKE--ELFEVSPKGFWLPDELAYDPIIPAILKDNKGYEYL 136
Db 85 YEPVLASIP---KEDRIEQIRLMKEMAKSICFGDARGVWMLTERVWQPELVKTLKESGIDYV 141
QY 137 FADGEAMLSAHLNSAIKPIKPLY-PHLIKAQRE-----KRFVYSYLLGLRELKRA 187
Db 142 IVD-----DYHFSAGLSKEELYWPYYTDEGGEVIAVFFIDEKLR---YLIFRPVDKV 192
QY 188 IKL---VFEKGVTLKAV--KDIEAVPVAVNTAVMLGIGRLPLMNPKKVASWI----- 236
Db 193 LEYLSHLDGDESKAVAVFHDGGEKFGIWPGETYEWY-----EKGWLRPEFDR 239
QY 237 --EDKNILLYGTDIE-----FIFYRDIAGYRMSVEGLLEVIDELNSCLP----- 281
Db 240 ISDEKINMLIYLEYLKYKPRGLVYLPYASY-----FEM-----SEWSLPAKQARLF 287
QY 282 ----SELKHSQ-----RELYLRTSSWAPKSLRIWREDEGN---ARLNMNLSYNNMRLGELAF 330
Db 288 VEFVNLKVKGIPEKYRVRVGGIW---KNF-FYKYPESNYMHKRLMVLMSKLYRNN---- 339
QY 331 AENS DARGWEPLPERRLDAFRAIYND--WRG 359
Db 340 -----PEARKYLLRAQCNDAYWHG 358

RESULT 3

US-07-894-212A-2
; Sequence 2, Application US/07894212A
; Patent No. 5366883
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UMORI, TAKASHI
; APPLICANT: MUKAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARB & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/07/894,212A
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 649 amino acids
; TYPE: amino acid

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS.

REFERENCE/DOCKET NUMBER: 82321
TELECOMMUNICATION INFORMATION:

SEQUENCE CHARACTERISTICS:

REFERENCE/DOCKET NUMBER: 62321

REFERENCE/DOCKET NUMBER: 62321

TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
TELEX: (25)3533
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-270-013B-2

Query Match 5.4%; Score 101; DB 1; Length 329;
Best Local Similarity 25.3%; Pred. No. 0.023;
Matches 39; Conservative 25; Mismatches 50; Indels 40; Gaps 7;
QY 10 LOYAEIPKSEIPK-----VIEKAYIPVETLKEEIPFGLNITGYTLKFLPKDIIDL 61
Db 130 MTYTVFESGFPKNRVIGSGVLDTR----FRFVAEE-----LNIS-----VKDVTGF 175
QY 62 VKGGIASDLIEIGTSYTHAI--LPLPLSRVEAQVQDRVKEELFEVSPKG--FWLPE 117
Db 176 VLGGHDDMVPLVRSYAGGIPLEKLIPLKDRDAIVERTRKGGGEIVNLLGNSAYYAPA 235
QY 118 LAYDPIIPAILKDN-----GYEYLF 137
Db 236 ASLVEMVEAILKDQRRILPAIAYLEGEYGEYI 269

RESULT 6
US-08-838-418-2
Sequence 2, Application US/08838418
Patent No. 5744342
GENERAL INFORMATION:

APPLICANT: Sogabe et al.
TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE
DEHYDROGENASE ACTIVITY
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:

ADDRESSEE: Leydig, Voit & Mayer, Ltd.
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60601-6780

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/838,418
FILING DATE: 17-MAR-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/270,013
FILING DATE: 01-JUL-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 164701/1993

FILING DATE: 02-JUL-1993

ATTORNEY/AGENT INFORMATION:

NAME: Hoover, Allen E.

REGISTRATION NUMBER: 37354

REFERENCE/DOCKET NUMBER: 78339

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 616-5600

TELEFAX: (312) 616-5700

TELEX: (25)3533

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 329 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-838-418-2

Query Match 5.4%; Score 101; DB 1; Length 329;
Best Local Similarity 25.3%; Pred. No. 0.023;
Matches 39; Conservative 25; Mismatches 50; Indels 40; Gaps 7;
QY 10 LOYAEIPKSEIPK-----VIEKAYIPVETLKEEIPFGLNITGYTLKFLPKDIIDL 61
Db 130 MTYTVFESGFPKNRVIGSGVLDTR----FRFVAEE-----LNIS-----VKDVTGF 175
QY 62 VKGGIASDLIEIGTSYTHAI--LPLPLSRVEAQVQDRVKEELFEVSPKG--FWLPE 117
Db 176 VLGGHDDMVPLVRSYAGGIPLEKLIPLKDRDAIVERTRKGGGEIVNLLGNSAYYAPA 235
QY 118 LAYDPIIPAILKDN-----GYEYLF 137
Db 236 ASLVEMVEAILKDQRRILPAIAYLEGEYGEYI 269

RESULT 7
US-09-134-001C-3678
Sequence 3678, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3678
LENGTH: 367
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3678

Query Match 5.0%; Score 94.5; DB 4; Length 367;
Best Local Similarity 18.7%; Pred. No. 0.14;
Matches 56; Conservative 52; Mismatches 118; Indels 73; Gaps 12;

QY 4 LVFPGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIIDLVK 63

Db 68 LIFKG-----VKRIVEDGYGSIIRKLIQNNI--NLIALHTNLDVNPKNVNRMLA 114

QY 64 GGIASDLIEIGT--SYTHAILPLPLSRVEAQVQDRVKEELFEVSPKGFWLPELAYD 121

Db 115 DQIGLENISMINTNSYYKVVQTFIPKNYIE-----DFKDSLNL-----154

QY 122 PIIPAILKDNQYELFADGEAMLFSAHLSAIPKIPPLPHLIKAQREKFRVYSYLLGL 181

Db 155 -----GLAKEONYEYCFESEG-----KQGFPGVDASPYICKLDS-----IEV---194

QY 182 RELRKAIKLVF-----EGKVTLLKAVKDIE--AVPV--WVAVNTAVMLGICRLPLMNPVKV 232

Db 195 ----DEIKLEFMKDNLEITKRAILDNHPYETPVDFIKMKNESEYGLIGQLAQWT 250

QY 233 ASWIEDKDNILLYGTDFIEFGYRDIAGYRMSVEGLLEV-----IDENSELCLPSELKH 286

Db 251 LDEFSEYAKQLNIPSVRYTQGHDSPIKVAIIIGSGIGFEYKASQLGADVFTGDIKH 309

RESULT 8
US-09-693-146-4
Sequence 4, Application US/09693146
Patent No. 6413758
GENERAL INFORMATION:
APPLICANT: Xu, Shuang-yong

Query Match 4.8%; Score 90.5; DB 4; Length 406;
Best Local Similarity 19.2%; Pred. No. 0.43;
Matches 64; Conservative 52; Mismatches 101; Indels 117; Gaps 13;

—

—

QY 173 -----RYSYLLGLREL--KKAIVLVEGKVTLKAVKDIEAVPVWVAIVTAVMLGIGR 223
Db 1196 ERIPAVGRNNQIYVSEKTAARRNAQVVF-----LRAISHTPLGLTTFSGARRALLQGLDE 1250
QY 224 LPLMPPKKVASWIEDKNILLYG---TDI-----EFIGYRDIAGYRMSVEGLLEVID 272
Db 1251 LERAQANSKVS-VQSSRIYHLSPQSDATPEETAKFEPEG-----VJD 1293
QY 273 ELNSELCLPSELKHSRELYLRTSSWAPDKSLRWREDE-GNAR-----LNMLSYNMRGE 326
Db 1294 KLKSLR-----AQLTKLRVDE--IETKVRVTVQDEGSPRVVVRVLVASSMOGE 1341

RESULT 13

US-09-134-001C-3428
; Sequence 3428, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3428
; LENGTH: 1154
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-3428

Query Match 4.7%; Score 88; DB 4; Length 1154;
Best Local Similarity 20.3%; Pred. No. 4.3;
Matches 57; Conservative 44; Mismatches 88; Indels 92; Gaps 15;

QY 107 EVSPKGFPLPELAYDPIIPAILKNDYELFADGEAMLSAHLNSAIKPIKPYPLHLKA 166
Db 391 EISP-----YDLSL---LVKLSHAVSFQAEEKMERSLRMRIRGVKTNIPPLINV 439
QY 167 QREKRRFYISYLLGLRE-----LRKATK-LVFEGKVTLKAVKDIEAVPVWVAVN 214
Db 440 MRNDKFRSGDYTKFTEETPELFDIAPTLDRGKTLEYIGNVTINGFPNVEKRPKEYES 499
QY 215 TAVMLGIGRLPLMPKK-----VASWIEDKNILLYGTDIEF-IGYR 255
Db 500 T-----KIPKISQKKINQFGTKOILEQHGPTGVTNWVREQEDVLI--TDTTFRDAHQ 550
QY 256 DIAGYRMSVEGLLEVIDELNSELCLPSELKHSRELYLRTSSWAPDK-SLRIRWREDEGNA 314
Db 551 SLIATVRTRKDMNMIAS-----KTAEVFKDSFSLMW-----GGA 585
QY 315 RLNLMSYNMRGELAENSADARGWEPLPERLLDAFRAYN 355
Db 586 TFD-VAYN-----FLKENP-----WERLERLR-----KAIPN 611

RESULT 14

US-08-248-021A-2
; Sequence 2, Application US/08248021A
; Patent No. 5648240
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Jonsson, Klas
; APPLICANT: Patti, Joseph M.
; APPLICANT: Gurusiddappa, Sivashankarappa
; TITLE OF INVENTION: MHC II ANALOG FROM STAPHYLOCOCCUS AUREUS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/248,021A
FILING DATE: 24-MAY-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: TANK:155
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 689 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-248-021A-2

Query Match 4.7%; Score 87.5; DB 1; Length 689;
Best Local Similarity 18.8%; Pred. No. 2.1;
Matches 51; Conservative 38; Mismatches 79; Indels 103; Gaps 11;
QY 33 ETLKEEIPFGNITGYTLKFLP-----KDIIDLKVGSIAD----- 69
Db 264 EAKVNNQVPYSINLTNQSNLAFSNKPWTNKLTTKVKSVLKSDRGVSRDLKHA 323
QY 70 -----LIEIGTSYTHAILPLPLSRVEAQVQDRVEKEEL---FEVSPK 111
Db 324 KKAYTVYFKNGKRVHLSNIYTNANLVHAKDVRRIEVTATVSKVRAERYVPTIAN 383
QY 112 GFWLPEL-----AYDPI---IPAILK-DNG-----YELFADGEAM 143
Db 384 GASNPDLSDLKFTGDSRVSYSDIKKKVSVLKHDRGIGERELKYAEKATYTVHFKNGTKK 443
QY 144 LFSALHNSAIKPIKPYPLHLK-----AQREKFRYISY----- 177
Db 444 VI---NLNSNISLNLLYVKDIKNIDIVKTAGAKVYSYVPTIANGTTTPIASKLKLS 501
QY 178 ---LLGLRELKRAIKLVFEGKVTLKAVKDIE 205
Db 502 NKOLIGYQDLNKKVSVLKH-----RGINDIE 529

RESULT 15

US-09-141-135-2
; Sequence 2, Application US/09141135
; Patent No. 5981729
; GENERAL INFORMATION:
; APPLICANT: CHUN, Jong Yoon
; APPLICANT: LEE, Yong Hun
; TITLE OF INVENTION: Transcription Factor Gene Induced by Water Deficit and Abscisi
; TITLE OF INVENTION: Acid Isolated from Arabidopsis thaliana
; FILE REFERENCE: 1942/31
; CURRENT APPLICATION NUMBER: US/09/141,135
; CURRENT FILING DATE: 1998-08-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 2
; LENGTH: 235
; TYPE: PRN
; ORGANISM: Arabidopsis thaliana

US-09-141-135-2

```

Query Match      4.68; Score 87; DB 2; Length 235;
Best Local Similarity 21.08; Pred. No. 0.42;
Matches 49; Conservative 37; Mismatches 71; Indels 76; Gaps 11;

QY 165 KAQREKRPRIYSYLLGLRELRLKAIKLVFECKVFLKAVKDIEAVPVWVAVNTAVMLGIGRL 224
      |: :||| |: :||| |: :||| |: :||| |: :||| |: :||| |: :|||
Db 27 KSNQKREN-----EEOIKSLLEIFSETLEPRKKVQ-----VARELG----- 65

QY 225 PLMNPKKVASWIEDK-----NTLYGTDIEFIGYRTDIAG-----YRMSVEGLL 268
      |: :||| |: :||| |: :||| |: :||| |: :||| |: :||| |: :|||
Db 66 --LQPRQMTWIFQNKRAWRKTKOLEKYNTLR-----ANTNNILASOFETMKKEKOSLV 116

QY 269 EVIDENSELCPSELKXH-----SGRELYLRTSSWAPD-KSLRIWREDSGNARLNLMSYN- 322
      |: :||| |: :||| |: :||| |: :||| |: :||| |: :||| |: :|||
Db 117 SELQLRLNEMQRPKEKHCCGDOGLAISSTESHGSKSEPEGRLDGSLCNDGDYNN 176

QY 323 -----MRG-----ELAFIAENSDARGVEPIPERLDAFRAIYNDWGENGE 363
      |: :||| |: :||| |: :||| |: :||| |: :||| |: :||| |: :|||
Db 177 NIKTEYFRQVQGETHELMNTVEKAD-----DSCLTSSNNWGCGFNDS 217

```

Search completed: November 18, 2002, 07:26:22
Job time : 23 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002, CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 07:24:24 : Search time 34 Seconds
(without alignments)
2205.916 Million cell updates/sec

Title: US-09-619-032-4
Perfect score: 1880
Sequence: 1 MRALVFHGNLQYAEIPKSEI.....RRLDAFRALYNDWRGENGP 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1544	82.1	364	17 Q9HHB5	Q9HHB5 pyrococcus
2	1506	80.1	364	17 Q58106	Q58106 pyrococcus
3	167	8.9	529	16 Q8YU2	Q8YU2 anabaena sp
4	161.5	8.6	529	16 P74630	P74630 synechocyst
5	152.5	8.1	378	17 Q97BM4	Q97BM4 thermoplasm
6	151	8.0	527	16 Q97GF3	Q97GF3 clostridium
7	142.5	7.6	443	17 Q973H0	Q973H0 sulfolobus
8	140.5	7.5	895	17 Q972N0	Q972N0 sulfolobus
9	135	7.2	632	17 Q8U136	Q8U136 pyrococcus
10	129.5	6.9	560	17 Q50094	Q50094 pyrococcus
11	127	6.8	923	16 Q9KD04	Q9KD04 bacillus ha
12	126	6.7	1362	17 Q9V294	Q9V294 pyrococcus
13	124	6.6	471	17 Q8ZXX1	Q8ZXX1 pyrobaculum
14	123.5	6.6	744	16 Q8YXA5	Q8YXA5 anabaena sp
15	122.5	6.5	357	17 Q9HL91	Q9HL91 thermoplasm
16	119.5	6.4	457	17 Q8ZT57	Q8ZT57 pyrobaculum

17	119	6.3	602	17 Q9VOM7	Q9VOM7 pyrococcus
18	115.5	6.1	324	17 Q30246	Q30246 archaeoglob
19	111	5.9	526	16 Q83377	Q83377 treponema p
20	109.5	5.8	447	17 Q97ZD2	Q97ZD2 sulfolobus
21	109	5.8	1069	16 Q8YI59	Q8YI59 anabaena sp
22	108	5.7	239	17 Q8TXM0	Q8TXM0 methanopyru
23	108	5.7	1355	17 Q8TRA6	Q8TRA6 methanosarc
24	107.5	5.7	902	17 Q97VY0	Q97VY0 sulfolobus
25	106	5.6	378	17 Q8TIT8	Q8TIT8 methanosarc
26	104	5.5	526	16 Q53278	Q53278 mycobacteri
27	104	5.5	546	16 Q8YVQ5	Q8YVQ5 anabaena sp
28	103.5	5.5	396	17 Q8TIT9	Q8TIT9 methanosarc
29	103	5.5	704	10 Q9ZVN2	Q9ZVN2 arabidopsis
30	103	5.5	754	13 Q13131	Q13131 oncorhynch
31	103	5.5	1150	16 Q9QUY8	Q9QUY8 staphylococ
32	102.5	5.5	899	17 Q8TNC5	Q8TNC5 methanosarc
33	102	5.4	592	16 Q8R5U6	Q8R5U6 thermoaer
34	102	5.4	754	13 Q13132	Q13132 oncorhynch
35	102	5.4	816	17 Q8ZUB8	Q8ZUB8 pyrobaculum
36	101.5	5.4	438	17 Q8U003	Q8U003 pyrococcus
37	101.5	5.4	684	10 Q9M2S4	Q9M2S4 arabidopsis
38	101	5.4	4845	11 Q88738	Q88738 mus musculu
39	100.5	5.3	266	16 Q97M42	Q97M42 clostridium
40	100.5	5.3	458	10 Q9MBD0	Q9MBD0 pyrus pyrif
41	100.5	5.3	619	16 Q83182	Q83182 treponema p
42	100	5.3	929	5 Q9VR32	Q9VR32 drosophila
43	99.5	5.3	1181	10 Q04650	Q04650 arabidopsis
44	99	5.3	393	5 Q8T2D0	Q8T2D0 dictyosteli
45	99	5.3	503	5 Q17585	Q17585 caenorhabdi

ALIGNMENTS

RESULT 1

Q9HHB5 PRELIMINARY; PRT; 364 AA.
AC Q9HHB5;
DT 01-MAR-2001 (TREMREL. 16, Created)
DT 01-MAR-2001 (TREMREL. 16, last sequence update)
DE 01-JUN-2002 (TREMREL. 21, last annotation update)
DE Alpha-galactosidase (Hydrolase).
GN GALA OR PF0444.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 3638;
RA Verhees C.H.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VCI / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF195244; AAG28455.1; -
DR EMBL; AE010167; AAL80568.1; -
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 1.
KW Complete proteome.
SQ SEQUENCE 364 AA; 41545 MW; 3E8E0AF5BDCCF2A5 CRC64;

Query Match 82.1%; Score 1544; DB 17; Length 364;
Best Local Similarity 79.9%; Pred. No. 5.4e-115;
Matches 290; Conservative 36; Mismatches 37; Indels 0; Gaps 0;
QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVTETLKEEIPGLNITCYTLKLPKDIID 60
DB 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVTETLKEEIPGLNITCYTLKLPKDIID 60

```

QY 61 LVKGIASDLIEITGTSYTHAILPLPLSRVQAQVQDRREVKEBELFEVSPKGFWLPDELAY 120
Db 61 LIKEGIESGLIEITGTSYTHAILPLPLSRVQAQVQDRREVKEBELFEVSPKGFWLPDELAY 120
QY 121 DPIIPAILKONGEYELFADGEMALFSAHLNSAIKPIKIPLYPHLLKAQREKFRFVSYLLG 180
Db 121 DPIIPAILRDNNYELFADGEMALFSAHLNSAIKPIKIPLYPHLLKAQREKFRFVSYLLG 180
QY 181 LRELKAIKLVFEGKVTLLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
Db 181 LRELKAIKLVFEGKVTLLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
QY 241 NLLYGTDIEFGYDIAGYRMSVEGLLEVIDELNSELCLPSELKHSRELYLRTSSWAP 300
Db 241 EILLYGTDIEFGYDIAGYRMSVEGLLEVIDELNSELCLPSELKHSRELYLRTSSWAP 300
QY 301 DKSRLRWDEGNARLNMLTSYMDGELAFLAENSDARGWEPPLPERRLDAPRAIYNDWRGE 360
Db 301 DKSRLRWDEGNARLNMLTSYMDGELAFLAENSDARGWEPPLPERRLDAPRAIYNDWRGE 360
QY 361 NGE 363
Db 361 NGK 363

RESULT 2
O58106 PRELIMINARY; PRT; 364 AA.
AC O58106;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Hypothetical protein PH0368.
GN PH0368.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamizawa M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000002; BAA29442.1; -
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 364 AA; 41755 MW; 7B4B36AB4A975BAD CRC64;

Query Match 80.1%; Score 1506; DB 17; Length 364;
Best Local Similarity 76.6%; Pred. No. 5.8e-112;
Matches 278; Conservative 47; Mismatches 38; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAFIPKSEIPKVEKAYIPVETLKEEIPFGNITGYTLKPLPKDID 60
Db 1 MRALVFHGNLQYAFIPKSEIPKVEKAYIPVETLKEEIPFGNITGYTLKPLPKDID 60
QY 61 LVKGIASDLIEITGTSYTHAILPLPLSRVQAQVQDRREVKEBELFEVSPKGFWLPDELAY 120
Db 61 LIKEGIESGLIEITGTSYTHAILPLPLSRVQAQVQDRREVKEBELFEVSPKGFWLPDELAY 120
QY 121 DPIIPAILKONGEYELFADGEMALFSAHLNSAIKPIKIPLYPHLLKAQREKFRFVSYLLG 180
Db 121 DPIIPAILRDNNYELFADGEMALFSAHLNSAIKPIKIPLYPHLLKAQREKFRFVSYLLG 180

```

```

QY 181 LRELKAIKLVFEGKVTLLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
Db 181 LRELKAIKLVFEGKVTLLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
QY 241 NLLYGTDIEFGYDIAGYRMSVEGLLEVIDELNSELCLPSELKHSRELYLRTSSWAP 300
Db 241 EILLYGTDIEFGYDIAGYRMSVEGLLEVIDELNSELCLPSELKHSRELYLRTSSWAP 300
QY 301 DKSRLRWDEGNARLNMLTSYMDGELAFLAENSDARGWEPPLPERRLDAPRAIYNDWRGE 360
Db 301 DKSRLRWDEGNARLNMLTSYMDGELAFLAENSDARGWEPPLPERRLDAPRAIYNDWRGE 360
QY 361 NGE 363
Db 361 NGK 363

RESULT 3
O8YUA2 PRELIMINARY; PRT; 529 AA.
AC O8YUA2;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein Alr2450.
GN Alr2450.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003589; BAB74149.1; -
DR InterPro; IPR003803; DUF200.
DR Pfam; PF02651; DUF200; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 529 AA; 62103 MW; 6AB10681738C8805 CRC64;

Query Match 8.9%; Score 167; DB 16; Length 529;
Best Local Similarity 21.0%; Pred. No. 6.3e-05;
Matches 104; Conservative 55; Mismatches 162; Indels 174; Gaps 21;

QY 3 ALVTHGNLQYAEIPKSEIPKVEKAYIPVETLKEEIPFGNITGYTLKPLPKDID 44
Db 7 ALVTHGNLQYAEIPKSEIPKVEKAYIPVETLKEEIPFGNITGYTLKPLPKDID 44
QY 45 -----NITGYTLKPLPKDID----- 56
Db 67 VSMRLDPLQLQERYDAHLAQLAELEIEGERRNAQNGHLRYLAHYATEFNEARQMWERYNG 126
QY 57 DIIDLVLKGIASDLIEITGTSYTHAILPLPLSRVQAQVQDRREVKEBELFEVSPKGFW 114
Db 127 DLVTAFAKQFQDSNNLEITTCGATHGYLPLMKMYQAVVQAQVQACEHYEEFGRRPKGIW 186
QY 115 LPELAYDPIIPAILKONGEYELFADGEMALFSAHLNSAIKPIKPL----- 159
Db 187 LPELAYDPIIPAILKONGEYELFADGEMALFSAHLNSAIKPIKPL----- 159
QY 160 -----YPHLLKAQR-----EKFRFVSYLLGLRELKAIKLVFEGKVT----- 197
Db 246 SQQVWSSEVGPAAEYEFYKDLGWAEAYEYIKPYIMPNGORKNTGKTYH-KITGRGLG 304
QY 198 --LKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKDNL--LYGT----- 247
Db 305 LSDKALYD---PYWAKEKA-----EHAANFMYNRQAEHLGLIMQRPPI 347

```

Db 363 EGPWFIDYLPKSWDFDQTEMTHLADYLRGNPHQQVCPRPSQSWGKGFHEWLNDDTNA 422

QY 261 -----RMSVEGLLEVIDELANSELCPSELKSHGRELYL-RTSSWA 299

Db 423 WIYPHLKAAERMTELSHREAVDELEK-----ALNQAARELLLAQSSDWA 468

RESULT 5

Q97BM4 PRELIMINARY; PRT; 378 AA.

AC Q97BM4;

DT 01-OCT-2001 (TrEMBLrel. 18, Created)

DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Alpha-amylase.

GN TV0431 OR TVG0421416.

OS Thermoplasma volcanium.

OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;

OC Thermoplasmataceae; Thermoplasma.

OX NCBI_TaxID=50339;

RP [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GSS1 / DSM 4299 / JCM 9571;

RC MEDLINE=20570466; PubMed=11121031;

RX Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,

RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,

RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;

RT "Archaeal adaptation to higher temperatures revealed by genomic

RT sequence of Thermoplasma volcanium."

EL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).

DR EMBL; AP000992; BAB59573.1; -

DR InterPro; IPR004300; Glyco_hydro_57.

DR Pfam; PF03065; Glyco_hydro_57; 1.

KW Complete proteome.

SQ SEQUENCE 378 AA; 44957 MW; 49FCFCF63BEE6B3A7 CRC64;

Query Match 8.1%; Score 152.5; DB 17; Length 378;

Best Local Similarity 21.4%; Pred. No. 0.00056;

Matches 67; Conservative 48; Mismatches 123; Indels 75; Gaps

QY 22 KVIEKAYIPVETLIKKEIPEPLNITG-----YTLKFLPKDIDLVKGGIASDLIEI 73

Db 42 RVAERSYIPATRNMEYGIKSFISITGAVTEQALMYNTK-----VIDAIDYVKSLCEM 96

QY 74 IGTSYTHAILPLPLSRVQAQVQDRVKELFVSPKFWLPCLAYDPIIPLKDNQY 133

Db 97 LSETYTHSLAIWNYDEKROVDHRLMKRIFNVVSPKVFRTNELIYDDRIAEMVKRMGF 156

QY 134 EYLFADGEMALFSAHLNSAIKPIKPLPHLIKAQREKFRFYIS---YLLGLRELKRAIKL 190

Db 157 TSIIITEG-----TDSIVKDHSPTY-----RYASPSGLNLYLRNVMSDNI 196

QY 191 VFEQKVTLKAVKDIEAVPVVAVTAVMLGTGRPLPNPKKVASWIEDKNILLYGTDIE 250

Db 197 SFRFSNT--KWKDP-----LTADKYAKWIINESEGDVWN---L 229

QY 251 FGYEDIAGYRMSVGEGLLEVIDEL-----NSELCLPSELKSHGRELYL--RTSSWA 299

Db 230 FMDYETFGEHOTQETGIDFMKYLPIVYFRDYGIEITITISEAKRHVRKDVLSIPETISWA 289

QY 300 PD-KSLRIWREDE 311

Db 290 DTRRDLRAWLENE 302

RESULT 6

Q97GF3 PRELIMINARY; PRT; 527 AA.

AC Q97GF3;

DT 01-OCT-2001 (TrEMBLrel. 18, Created)

DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Uncharacterized conserved protein, related to
GN alpha-amylase/alpha-mannosidase.

OS CAC2414.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RA MEDLINE=21359325; PubMed=11466286;
RX Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AEO07742; AAK80369.1; -
DR Interpro: IPR003803; DUF200.
DR Pfam: PF02651; DUF200; 1.
KW Complete proteome.
SQ SEQUENCE 527 AA; 61485 MW; 03706E16907929C6 CRC64;

Query Match 8.0%; Score 151; DB 16; Length 527;
Best Local Similarity 18.9%; Pred. No. 0.0012;
Matches 92; Conservative 63; Mismatches 119; Indels 212; Gaps 20;

QY 5 VFGNLOAYAIKPKSE-----IPKIEKAYIPVIE--TLIKEIPFG----- 43
DB 9 VLHSHMPFVRHPTROSLEERWLFEMSECYPLIEVYNLLKNKIKFRMTWSITPPLMS 68
QY 44 -----LNTGYTLKFLPKDII-----DL 61
DB 69 MLQDEVLSRYLNLAKTKTIELSEKEILRTKNNPEENKVALFYNKRAENTLKIEYKYNL 128
QY 62 VKGIASD---LLEIGTSYTHAILPLPLSR--VEAQVQDREKVEELFEVSPKGFWLP 116
DB 129 INAFKRYDLGCVETITCAATHALLPLILNQAVKRAQIATGVQSYNTMGHEPNGLWP 188
QY 117 ELAYDPIPALKNDGYEVLFDAGEAMLFSAHLNSAIKIPKLYPHILKAQREKFRYIS 176
DB 189 ECAYTGIDNLSEFGIKFIISGKAIDYASP-----KMY----- 224
QY 177 YLLGLRELKAIKLVEFGKVTLKAVKDIEAVPVWAVNTAVMLGIGRLPLMNPKKVASWI 236
DB 225 -----GTNT-----PIAPSGVCAFG 240
QY 237 EDKDNILLYCTDIEFG-----YRDIAGYRMSVGLLEVIDELNSCLPSELKH 286
DB 241 RMDS--SYQVMSDFMGYPGDFNYREFYRDI-GFELPMYIKPYINE--NGRIDTGRKY 295
QY 287 -----SGRE-LYLRSS-----WAPDKSLRIWREDEGN--ARLNM----- 318
DB 296 YKIGNSGKGYNRENAMKWEHSHAFSCRHQDINAAAANMDKPPITPCPYDTLYG 355
QY 319 -----LSYNNRGELAFLAENSARGWEPLPERRLDFAFRYINDW 357
DB 356 HWMFEGPDFINAFIRKSAEDWTSYELITPTVEYLKNNMVQCSSPSP-----SSW 404
QY 358 RGENGE 363
DB 405 -GENGD 409

RESULT 7
Q973T0 PRELIMINARY; PRT; 443 AA.
AC Q973T0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein ST0817.

GN ST0817.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
Crenarchaeon, Sulfolobus tokodaii strain 7";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000983; BAB65830.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 443 AA; 53218 MW; 497087C264BC03C4 CRC64;

Query Match 7.68%; Score 142.5; DB 17; Length 443;
Best Local Similarity 23.0%; Pred. No. 0.0044;
Matches 79; Conservative 41; Mismatches 134; Indels 89; Gaps 15;

QY 22 KVIEKAYIPVETLKEEIPFGLNITGYTLKFL-----PKDIIDLKVGIA 67
DB 46 RVKKCYIPATRIILL-EEIENG-ESEGYDFRFFSLSGTFIEQAQKGVLELQOLSY 103
QY 68 SDLIEIGTSYTHAILPL-PLSRVEAQVQDREKVEELFEVSPKGFWLPAYDPIIPA 126
DB 104 THKVEFLSQTYVHSITSLWEDLTWKKEQVQMHKNLRDYFGQEPVTFENTELLTPRIVK 163
QY 127 ILKNGEYELFADGEAMLFSAHLNSAIKIPKLYPHILKAQREKFRYISYLLGLRELK 186
DB 164 EIEKLFKVVITEGK-----ESLLKKGSP-----NRVYRIRDTKLSLLLRVRL 207
QY 187 AIKLVFEKGVTLKAVKDIEAVPVWAVNTAVMLGIGRLPLMNPKKVASWIEDKDNILLYG 246
DB 208 SDDIAFR-----FSNPKW-----DQPL-TASKFADWVSWSEGEI--- 241
QY 247 TDIEFGYRDIAGYRMSVGLLEVIDELNSCLPSELKHSG-----RELYLR----- 294
DB 242 -GLIFVDYETFEHHWPESGILDFL-----RWLPRELHRRGVFEKLPREVYKYYEIV 294
QY 295 ---TSSWA---PDKS-----LRWREDEGNARLNMVSYNNRGE 326
DB 295 IDSTVSWADINKDESSWLGNIMQWAYDEMVRRTMLEAKEAGGE 337

RESULT 8
Q972N0 PRELIMINARY; PRT; 895 AA.
AC Q972N0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein ST1102.
GN ST1102.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,

Db	120	EINGDIIGKFRPEQEA	YLEIITSAATHGYLPLGRDEAIBGQIANAKTYEKYFQRRPR	179	
QY	112	GFWLP	PELAYDPI-----IPAILKONGVEYLFADGAMLFSAHLNSAIKP	155	
Db	180	GMWLP	PECAYPDGYWSPSTGEVWVRKGIHEFLKYGIDYFPFVESH-LIDEGPATSKYGE	238	
QY	156	IKP	-----LYPHLIK-----AQERKFRYISYLG-----LRELKRAIKLVFGK	195	
Db	239	ILPEKTKSTLRPYFLKNGI	AVFARNRETGIQWSADIGYPGDFWYRBFHK-----	289	
QY	196	VTLKAYKDI	AVPVAVTAVMLGIRLPLNPKK-----VASWIED--	238	
Db	290	---KAEK---	SGGOYWRVTKEIDLG-GKEPYV-PEKAMERVEEHARHFISLVSSLLLEBFK	342	
QY	239	---	KNILYGTGDI	EFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSEL--KHSGR--E	290
Db	343	KKEGEY	GIVVAPDYDTFLFGHWFEGVKW-LGKVLAEKLGIKTTTISEFLDNFDGKRYE	401	
QY	291	LYLRTSSW	-----APDKSLRWREDEGNARLNMLSYNMRGEL	327	
Db	402	IELPEGS	WGMFGTHYWNPEVETWPIIHLAEDRMVALASKYLGRLDTRTLDQLGRE	461	
QY	328	AFLAENS	DGRWE	340	
Db	462	LLLTSSD	---WQ	471	
RESULT 10					
ID	050094	PRELIMINARY;	PRT;	560 AA.	
AC	050094;				
DT	01-JUN-1998	(TReMBLrel. 06, Created)			
DT	01-JUN-1998	(TReMBLrel. 06, Last sequence update)			
DE	01-MAR-2002	(TReMBLrel. 20, Last annotation update)			
DE		Hypothetical protein PH1386.			
GN	PH1386.				
OS	Pyrococcus horikoshii.				
OC	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;				
OC	Pyrococcus.				
OC	NCBI_TaxID=53953;				
RP	[1]				
RN	SEQUENCE FROM N.A.				
RC	STRAIN=OT3;				
RC	MEDLINE=98344137; PubMed=9679194;				
RA	Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,				
RA	Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,				
RA	Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,				
RA	Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,				
RA	Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,				
RA	Masuchi Y., Shizuya H., Kikuchi H.;				
RT	*Complete sequence and gene organization of the genome of a hyper-				
RT	thermophilic archaeobacterium, Pyrococcus horikoshii OT3.*;				
EL	DNA Res. 5:55-76(1998).				
DR	EMBL; AP000006; BAA30492.1; -.				
DR	InterPro; IPR003803; DUF200.				
DR	Pfam; PF02651; DUF200; 1.				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE	560 AA; 65769 MW; 3F4CEP5267D167F8 CRC64;			
Query Match					
Best Local Similarity 25.6%; Score 129.5; DB 17; Length 560;					
Matches 60; Conservative 28; Mismatches 81; Indels 65; Gaps					
QY	14	EIPKSEIPKVI	KAYIPVIEFL-----IKKEIPFGLNITGYTLKFLPKDII	122	
Db	71	EYIKREEFKY	MRKLKSMEDLRFKDEKUREAIN---MIGYF-----KDI	122	
QY	64	GGAISL	-----EIIIGTSYTHAIPLLPLSR-VEAQVQRDR	114	
Db	123	GNILCKFREL	QDEGYEVVITSAATHGYLPLGRDEAIEAQLNGIKVYK	182	
QY	115	LPELAYDPI	-----IPAILKONGVEYLFADGAMLFSAHLNSAIKP	158	

```

DB 183 LPECAYPDGLWKSPSTGEVKWRKGLIEHFLKKGIEYFEVE-----SHLIDK-GPVSL 234
QY 159 LYPHLLKAOREFRYISYLLGLRELKRAIKLVFEKGVTLKAVKDIEAVPWVA 212
DB 235 RYGNILPA-KTKNSTLRYP-----LKNGIADFARNRET-----GIQVWSA 274

RESULT 11
Q9KD04
ID Q9KD04 PRELIMINARY; PRT; 923 AA.
AC Q9KD04;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Hypothetical protein BH1415.
GN BH1415.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans" and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AF001512; BAB05134.1; -.
DR InterPro; IPR003803; DUF200.
DR InterPro; IPR001296; Glycos_transf_1.
DR Pfam; PF02651; DUF200; 1.
DR Pfam; PF00534; Glycos_transf_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 923 AA; 106638 MW; 8C7B/D64F3603553 CRC64;

Query Match 6.8%; Score 127; DB 16; Length 923;
Best Local Similarity 23.8%; Pred. No. 0.22;
Matches 54; Conservative 31; Mismatches 70; Indels 72; Gaps 11;

QY 3 ALVHGNLOYAEIPKSE-----IPKVEKAYIPVIEILKEEIPFGLNI----- 46
DB 7 SLVLHAFLPYVRHQEDRLERLWFLEAMSETYIPLWALEKLPVKHVAVTISFTPPVMEML 66
QY 47 -----TGVT-----TLKFLPKD-----TIDLKVG-----G 65
DB 67 SDPLVQTRVLYNLHLENTQLLKEERTNDQRTQNLVQFYKQYKELKATFLQWRNLLIG 126
QY 66 IASDLIE-----IIGTSYTHAILPLPLSR-VEAQVQDRDREVKEELFEVSPKGFWLPPELA 119
DB 127 FRS-LWENEOCTLMTSAATHAFPPYKTKETRAQVRHGIACFEQHGKGLPGFWLPPECA 185
QY 120 YDPIIPAILKONGYLYLADGEMLF---SAHLSNAIKPIKPLY-PH 162
DB 186 FSPGVDRILFEEGIRYTFVDEHAVLTADTPHKGSS-----APIYSPH 228

RESULT 12
Q9V294
ID Q9V294 PRELIMINARY; PRT; 1362 AA.
AC Q9V294;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE AMYLOPULLULANASE.
GN PYO OR PAB0122.
OS Pyrococcus abyssii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;

Query Match 6.6%; Score 124; DB 17; Length 471;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssii genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248283; CAB49104.1; -.
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 2.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 1362 AA; 154578 MW; 277AFAB4E14860D1 CRC64;

Query Match 6.7%; Score 126; DB 17; Length 1362;
Best Local Similarity 21.8%; Pred. No. 0.46;
Matches 83; Conservative 56; Mismatches 142; Indels 100; Gaps 16;

QY 32 IETLIKEIPFLNITGYTLKFLPKDIDLVKGGIASDLIEIIGTSYTHAILPLP---- 87
DB 236 VETVLKHOM-WLLNHT-----FEEHEKINLLG---NGNVEVTVPYTHPIGILNDFGW 286
QY 88 LSRVEAQVQDRDREVKEELF---EVSPKGFWLPPELAIDPIIPAILKONGYLYLADG---E 141
DB 287 YEDFDAQVKKANELYKEYLGAGKVTPKGGWAESALNDKTLLELAENGWKWMTDQLVLE 346
QY 142 AMLFSAHLNSAIKP-----IKPLYPHLIKA 166
DB 347 KLGVPKTIESYKPPWVAQFGDKKILYLPFRNHDLSDRVGFYAGMNOYDAVKNFVEELKI 406
QY 167 QREKRFYISYLLGLRELKRAIKLVFEGKVTLKAV-KDIEAVPWWAVNTAV-----ML 219
DB 407 QKQYDGLSVVITLDGENPWEHYPPFDGKLFLEELYRQLEELQKGLIRTVTPSEVIEMF 466
QY 220 G-IGRLPLANPKKVASWIEDKNILLYGTDIEFIGYRDIAGYRMSVGLLEVIDELNSE 277
DB 467 GDKANKLTPKMKRKLDTEDNVNALLKAKTL-----GELYDMGVGTE----- 509
QY 278 LCLPSELKHSGRELYLRTSSWAPDKSLRIWREDEGNARLNMLSYNMRGELAFLAENSD-A 336
DB 510 -----EMQWPESWI-DGTLSTW---IGEPQENIAWYLYLARKALFENKDNV 553
QY 337 RGWEPLPERLDAFRAIYNW 357
DB 554 DKWNKAYE---LFRAPGSDW 571

RESULT 13
Q8ZXX1
ID Q8ZXX1 PRELIMINARY; PRT; 471 AA.
AC Q8ZXX1;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Hypothetical protein PAE1048.
GN PAE1048.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum."
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009801; AAL63225.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 471 AA; 53638 MW; B7B379D3E214FB04 CRC64;

Query Match 6.6%; Score 124; DB 17; Length 471;

```


QY 120 YDIPIAILKONGYEVLFADGEMALFSAHLNSA-IKPIKPLYPHLIKAKQREKFRFY----- 174
 Db 126 WEPELVKTLREAGIEVILD-----DYHMSAGLSKEELFWPYTENGGEAIVVFPIDE 179
 QY 175 -ISYLLGLRLKAKLVF-----EGKVLKAVKDIEAVPVWAVNTAVMLGIGRLPLM 227
 Db 180 KURLYLPFRVNETLEYLHSLADEDESKVAV-FHDDGEKFGAMPQTHELY----- 229
 QY 228 NPKKVASWI-----EDKNILLYGTDIEFIG-----YRDIAGY-RMSVEGL-- 267
 Db 230 -----ERGLKEFFDRISDDKINLMYS-----EYLSKFRPKGLVLYLPPIASIFEMSEWSLPA 282
 QY 268 -----LEVIDELNSLCLPSLSEKHSRELYLRTSSWAPDKSLRTWRDEGN---ARLNM 318
 Db 283 ROAKLFFEFIKKL-KELNLFERYR-----IFVRGGIW---KNF-LYKYPEGNYMHKRLM 332
 QY 319 LSVNMRGELAFLAENSARGWEPLPRLDAFRAYND--WRGENG 362
 Db 333 LSKLLRN-----PTARIFVLRAOCNDAYHGVFG 362

RESULT 2
 AMYL_DICTH
 ID AMYL_DICTH STANDARD; PRT; 685 AA.
 AC P09561;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Alpha-amylase 1 (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase).
 GN AMYA.
 OS Dictyoglomus thermophilum.
 OC Bacteria; Dictyoglomus group; Dictyoglomus.
 OX NCBI_TaxID=14;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H-6-12;
 RX MEDLINE=88225097; PubMed=2453362;
 RA Fukusumi S., Kamizono A., Horinouchi S., Beppu T.;
 RT "Cloning and nucleotide sequence of a heat-stable amylase gene from
 an anaerobic thermophile, Dictyoglomus thermophilum.";
 RL Eur. J. Biochem. 174:15-21(1988).
 CC -!- FUNCTION: THIS AMYLASE IS A HIGHLY LIQUEFYING-TYPE: OLIGOMERS
 APPEARED AT THE BEGINNING OF INCUBATION, FOLLOWED BY A GRADED
 DECREASE IN THE AMOUNTS OF MALTOTRIOSE, MALTOSE AND GLUCOSE IN
 PROLONGED INCUBATION. IT IS HIGHLY HEAT-STABLE; THE OPTIMUM
 TEMPERATURE FOR ITS ACTIVITY WAS FOUND TO BE AROUND 90 DEGREES
 CELSIUS, BUT DECREASED TO ONLY 70% ACTIVITY AFTER 1 HOUR. NO
 DECREASE OF ACTIVITY WAS OBSERVED WITHIN THE SAME TIME AT 80
 DEGREES CELSIUS.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 linkages in oligosaccharides and polysaccharides.
 CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; X07896; CAA30735.1; -
 DR PIR; S00628; ALDYAT.
 DR InterPro; IPR0004300; Glyco_hydro_57.
 DR Pfam; PF03065; Glyco_hydro_57; 1.
 KW Hydrolase; Glycosidase; Carbohydrate metabolism; Multigene family.
 FT. INIT_MET 0
 SQ SEQUENCE 685 AA; 81060 MW; C10941C8A508C404 CRC64;

Query Match

8.5%; Score 159.5; DB 1; Length 685;

Best Local Similarity 21.0%; Pred. No. 9.6e-05;
 Matches 85; Conservative 64; Mismatches 149; Indels 107; Gaps 19;

QY 8 GNLOYAEIPKSEIPKVIKAVIPVETLIKE-ETPFGLNITGYTLKPLK--DIIDLVK 63
 Db 17 GNDFV-----TERAVEMSYKPLINFFFKHPDPPINVHFGSGLLWLEKNHPEYFEKLK 70
 QY 64 GGIASDLIEIGTSYTHAILPLPLSRVQAQVORDREVEKEELFEVSPKGFELPDLAYDPI 123
 Db 71 IMAERGQIEFVSGFYEPILPIIPDKVOQIKLNKYIYDKFGQTPKGMWLAERVWEPH 130
 QY 124 IPAILKONGVEYLFADGEMALFSAHLNS-----AKPIKPLYPHLIK-A 166
 Db 131 LVKYIARAGIEYVVD--DAHFFSVGLKEEDLFGYLLMEEQYKLAVPISMKRLYLIPTFA 189
 QY 167 QREKFRFYISVLLGLRLKRAIKLVFEGKVTLKAVIDEAVPVWAVNTAVMLGIGRLPL 226
 Db 190 DPETITYYLDKFA--EDKSKIALFD-----DGEKFGLPDTYRTYV----- 230
 QY 227 MNPKKVASWIED-----KDNILL-----YGTDIEFIG--YRDIAGYRMSVEGLL--- 268
 Db 231 -----EEGWLETFVSKIKENFLNLTVPVNLTYMQRVKPKGRILYLPATASYREMMEWVLFPE 285
 QY 269 -----EVIDELNSLCLPSLSEKHSRELYLRTSSWAPDKSLRTWR-----DEGNARLN 317
 Db 286 AQKELEELVEKLTKE-----NLW---DKFSYVYKGGFWRNFLAKYDESN----- 326
 QY 318 MLSYNMRGELAFLAENSARGWEPLPRL--DAFRAYND--WRG 359
 Db 327 ---HMOKKMLYVNMKVDQSPNEEVEKEMEEVFGQANDAYWHG 367

RESULT 3
 AMYA_METJA
 ID AMYA_METJA STANDARD; PRT; 467 AA.
 AC Q59006;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative alpha-amylase (EC 3.2.1.1).
 GN M01611.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2651 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Rotodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 linkages in oligosaccharides and polysaccharides.
 CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; U67601; AAB99631.1; -

DR TIGR: MJ1611; -
DR InterPro: IPR004300; Glyco_hydro_57.
DR Pfam: PF03065; Glyco_hydro_57; 1.
KW Hypothetical protein; Hydrolase; Glycosidase; Carbohydrate metabolism;
SQ Complete proteome.
SEQUENCE 467 AA; 5558 MW; 40A6B1CDD4D967E CRC64;

Query Match 7.8%; Score 146.5; DB 1; Length 467;
Best Local Similarity 23.1%; Pred. No. 0.00058;
Matches 87; Conservative 66; Mismatches 129; Indels 95; Gaps 24;

QY 14 EIPKSEPKVIEKAYIPVETLIK--EETPFGLN--ITG-----YTLKFLPKD-IIDLKVG 64
DB 34 KLNKEVFNKANKCYPTNELLLELDEYDFKVNYSITGVFEQALEF--NDYVLDLPD 91
QY 65 GIASDLIEIGTSYTHAILPLPL-SRVEAQVORDREVKEELFEVSPKGFWLPCLAYDPI 123
DB 92 LKVTGNVELIAETVHSLTSLPTEDEFTEDIEHMKMYKEIFGPKAKVFRNTELYNNR 151
QY 124 IPAILKNGVEYLFADG-EMALFSAHLNSAIKPIKPLPHILKAQREKFRYIS-----YLL 179
DB 152 IAKIAKDLGPKAIFTEGIEKIL-----GWRSPNLYQSPDGMKI 190
QY 180 GLRELKAIKLVFEGKVTLKAVKDIEAVP-----VWVAVNTAVMLGIGRLPLMNPKV 232
DB 191 LLNRYLSDDIGFR-----FSARDWQYPLTADKYAIWLASTPGEVINI-----YMDYET 241
QY 233 AS--WIEDKDNILLYGTDI-EFTGYR--DIAGY-RMSVEGLLEVIDELNSLCLPSELKH 286
DB 242 GEHWKE-----TGIEFRLYPLPIETAKHEHLEVVNVSEVDRLEPR----- 283
QY 287 SGRLEYLR---TSSWA-PDKSLRIWREDEGNARLNLMLSYNMRGELA-FLAENSARGWEP 341
DB 284 --GEIYVHEEATISWADTSDVSAWL---GN-KMORISFEKLDIGKFKENSNIKL----- 333
QY 342 LPERLDAPRAIYNDWR 358
DB 334 ---KLNKFEIYMYK 347

RESULT 4
AMYA_PYRAB STANDARD; PRT; 655 AA.
AC Q9V238;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-amylase (EC 3.2.1.1).
GN AMYA OR PAB0118.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RA Hellig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
linkages in oligosaccharides and polysaccharides.
CC -1- PATHWAY: POLYSACCHARIDE DEGRADATION.
CC -1- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: AJ248283; CAB49100.1; -
DR InterPro: IPR004300; Glyco_hydro_57.
DR Pfam: PF03065; Glyco_hydro_57; 1.
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Complete proteome.
SQ SEQUENCE 655 AA; 77296 MW; 7F6F920B1A00EECE CRC64;

Query Match 7.5%; Score 141.5; DB 1; Length 655;
Best Local Similarity 21.9%; Pred. No. 0.0022;
Matches 88; Conservative 54; Mismatches 130; Indels 129; Gaps 21;

QY 25 EKAYIPVETLIKPEIP---FGLNITGYTLKFLPK---DIIDLKVGGIASDLIEIGTSY 78
DB 28 EKAYRPFEIL--EYPPNKAIVHISGILVEWLEENKPDYIDLKSLVRKQGVETVWAGF 85
QY 79 THAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPCLAYDPIIPAILKNGVEYLF 138
DB 86 YEPVLAALPKEDRLQEIYLLKEWAKKI-GYDAKGLMLTWRVWQPELVKTLREAGIEYVV 144
QY 139 DGEAMLFSAHLNSA-IKPIKPLPHILKAQRE-----KRFY-----ISYLL 179
DB 145 D-----DYHFMASGLSKDQLFWPYTDEGGEVITVFIDEKRLYLIFRPVDPKVISYL 198
QY 180 GLRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDK 239
DB 199 SLASEDESKVAVFH-----DDGEKFGIW-----PM-----TYEWVYEK 231
QY 240 D-----NLLYGTDI-----FIGYRDIAGY-----RMSVEG 266
DB 232 GWLREFFDRVSSDEAINIMLYSEYLOKPKGLVLYPLIASYFEMSEWSLPAQAKLFVE- 290
QY 267 LLEVIDELNSLCLPSELKHSRELYLRTSSWAPDKSLRIWREDEGN---ARLNLMLSYN 323
DB 291 FVEKIKELN-----MFEYRVFVGGIW---KNF-FYKYPEANYMHKRLMLSLRL 337
QY 324 RGEFLAENSARGWEPLEPERRLDAPRAIYND--WRGENG 362
DB 338 RDN-----PSARREVLRAQNDAYWHGVFG 362

RESULT 5
AMYA_PYRFU STANDARD; PRT; 648 AA.
AC P49067;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alpha-amylase (EC 3.2.1.1).
GN AMYA OR PF0272.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94043280; PubMed=8226990;
RA Laderman K.A., Asada K., Uemori T., Mukai H., Taguchi Y., Kato I.,
RA Anfinsen C.B.;
RT "Alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus
furiosus. Cloning and sequencing of the gene and expression in
Escherichia coli."
RL J. Biol. Chem. 268:24402-24407(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=94043279; PubMed=8226989;
RA Laderman K.A., Davis B.R., Krutzsch H.C., Lewis M.S., Griko Y.V.,
RA Privvalov P.L., Anfinsen C.B.;
RT "The purification and characterization of an extremely thermostable

RT alpha-amylase from the hyperthermophilic archaeobacterium *Pyrococcus furiosus*.";
 RL J. Biol. Chem. 268:24394-24401(1993).
 CC -!- FUNCTION: DISPLAYS A BROAD RANGE OF SUBSTRATE SPECIFICITY, WITH
 CC THE CAPACITY TO HYDROLYZE CARBOHYDRATES AS SIMPLE AS MALTOSE.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 CC linkages in oligosaccharides and polysaccharides.
 CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- MISCELLANEOUS: THE ISOELECTRIC POINT IS 4.3. THE ENZYME DISPLAYS
 CC OPTIMAL ACTIVITY, WITH SUBSTANTIAL THERMAL STABILITY, AT 100
 CC DEGREES CELSIUS, WITH THE ONSET OF ACTIVITY AT APPROXIMATELY 40
 CC DEGREES CELSIUS.
 CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L22346; AAA72035.1;
 DR EMBL; AF010151; AAL80396.1; ALT_INIT.
 DR InterPro; IPR004300; Glyco_hydro_57.
 DR Pfam; PF03065; Glyco_hydro_57; 1.
 KW Hydrolase; Glycosidase; Carbohydrate metabolism; Complete proteome.
 FT INIT MET 0
 SQ SEQUENCE 648 AA; 76178 MW; BF7A495F084E0FB1 CRC64;

Query Match 7.5%; Score 140.5; DB 1; Length 648;
 Best Local Similarity 22.8%; Pred. No. 0.0027;
 Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;
 QY 25 EKAYIPVETLKEEIP---FGLNITGYTLKFLP---KDIIDLKVGGSADLIEIGTSY 78
 DB 28 EKYWPFLLETL--EYPMKVAIHTSGPLIEWQDNRPEDYIDLSLVKRGQVELIVAGF 85
 QY 79 THAILPLPLSRVQAQVORDREVKE--ELFEVSPKGFWLPALAYDPIIPAILKNDGYEYL 136
 DB 86 YEPVLASIPKEDRIVQIEKLEKFAENLGYEA--RGVWLTERTVWQPELVKLSRAAGIDYVI 143
 QY 137 FADGEAMLFSAHLNSAIKPIKPLY--PHLIKAQRE-----KFRYISYLLGLRELKAI 187
 DB 143 IVD-----DYFMAGSLSKDELEWPPYTTEDGGEVIAVFPIDEXLR---YLIPRPVDKV 193
 QY 188 IKL---VFEGKVTLKAV--KDEAVPVVAVNTAVMLGIGRLPLMNPKKVASWI----- 236
 DB 194 LEYLHSLIDGDESKVAVFHDGGEKFGIWPYEWY-----EKGWLEFFEDR 240
 QY 237 ----EDKNILLYGTDIE-----FIGYRDIAGY-RMSVEGL-----LEVIDELNSELCLP 281
 DB 241 ISSDEKINILMYTELYEKPKRGLVPLPIASY-----FEM-----SEWSLPKAKQARLF 288
 QY 282 ----SELKHSQ-----RELVLRTSSWAPDKSLRIWREDEGN---ARLNMLSYNNMRGELAF 330
 DB 289 VEFVVELKVGIFEKYRVFVGGIW---KNF-FYKPSYNNMRKMLVSKAVRNN----- 340
 QY 331 AENSARGWEPLERRLDRAFRAYND--WRG 359
 DB 341 -----PEARKYLLRAQCNDAYWHG 359

RESULT 6
 MALQ_PYRKO
 ID MALQ_PYRKO STANDARD; PRT; 653 AA.
 AC 032450;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)
 DE (Disproportionating enzyme) (D-enzyme).

OS *Pyrococcus kodakaraensis*.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Thermococcus.
 OX NCBI_TaxID=69014;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KOD1;
 RA Tachibana Y., Fujiwara S., Takagi M., Imanaka T.;
 RT "Cloning and expression of the 4-alpha-glucanotransferase gene from
 RT the hyperthermophilic archaeon *Pyrococcus* sp. KOD1, and
 RT characterization of the enzyme.";
 RL J. Ferment. Bioeng. 83:540-548(1997).
 CC -!- CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucan
 CC to a new 4-position in an acceptor, which may be glucose or (1,4)-
 CC alpha-D-glucan.
 CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D87907; BAA22062.1;
 DR InterPro; IPR004300; Glyco_hydro_57.
 DR Pfam; PF03065; Glyco_hydro_57; 1.
 KW Transferase; Glycosyltransferase; Carbohydrate metabolism.
 SQ SEQUENCE 653 AA; 76664 MW; B0C4695613F29219 CRC64;

Query Match 6.5%; Score 122.5; DB 1; Length 653;
 Best Local Similarity 21.4%; Pred. No. 0.067;
 Matches 83; Conservative 65; Mismatches 138; Indels 101; Gaps 22;
 QY 25 EKAYIPVETLKEEIP---FGLNITGYTLKFL---PKDIIDLKVGGSADLIEIGTSY 78
 DB 28 ERSTRPPEMTL--EYPMKVAIVHYSGLPLEWIRDNKPEHLDDLRLSLVRKQLEIVVAGF 85
 QY 79 THAILPLPLSRVQAQVORDREVKEEL--FEVSPKGFWLPALAYDPIIPAILKNDGYEYL 137
 DB 86 YEPVLASIPKEDRIVQIEKLEKFAENLGYEA--RGVWLTERTVWQPELVKLSRAAGIDYVI 143
 QY 138 ADGEAMLFSAHLNSA--IKPIKPLYPHLIKAQRE-----KFRYISYLLGLRELKAI 188
 DB 144 VD-----DYFMAGSLSKDELEWPPYTTEDGGEVITVFPIDEXLR---YLIPRPVDKTL 194
 QY 189 KLVF---EGKVTLKAV--KDEAVPVVAVNTAVMLGIGRLPLMNPKKVASWI----- 236
 DB 195 EYLHSLIDGDESKVAVFHDGGEKFGVWPGYEWY-----EKGWLEFFEDR 241
 QY 237 --EDKNILLYGTDIE-----FIGYRDIAGY-RMSVEGL-----LEVIDELNSELCL 280
 DB 242 SSDERINILMYSEYLRFRPRGLVPLPIASYFEMSEWSLPARQAKLFVEVEELK----- 296
 QY 281 PSELKHSGRELYLRTSSWAPDKSLRIWREDEGN---ARLNMLSYNNMRGELAFIAENSAR 337
 DB 297 -KENKEDRVRFVGGIW---KNF-EFKYPSYNNMRKMLVSKAVRNN----- 340
 QY 338 GWEPLPERRLDAFRAYND--WRGENG 362
 DB 341 -----PEAREFILLRAQCNDAYWHGVEG 362

RESULT 7
 MALQ_THELI
 ID MALQ_THELI STANDARD; PRT; 659 AA.
 AC 032462;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)
 DE (Disproportionating enzyme) (D-enzyme).

[illegible]

RESULT 8	MDH_BACSU	STANDARD;	PRT;	311 AA.
ID	MDH_BACSU			
AC	P49814;			
DT	01-OCT-1996 (Rel. 34, Created)			
DD	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Malate dehydrogenase [EC 1.1.1.37] (Vegetative protein 69) (VEG69).			
GN	CITH OR MDH.			
OS	Bacillus subtilis.			
OX	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
NCBI	NCBI_TaxID=1423;			
[1]	SEQUENCE FROM N.A.			
RP	STRAIN=168 / SMY;			
RC	MEDLINE=96134995; PubMed=8550482;			
RX	Jin S., de Jesus-Berrios M., Sonenshein A.L.; "			
RT	"A Bacillus subtilis malate dehydrogenase gene. "			
RL	J. Bacteriol. 178:560-563(1996).			
[2]	SEQUENCE FROM N.A.			
RP	STRAIN=168;			
RC	MEDLINE=98048467; PubMed=9387221;			
RX	Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;			
RT	"sequencing and functional annotation of the Bacillus subtilis genes			
RL	in the 200 kb rnb-dnaB region. "			
[3]	Microbiology 143:3431-3441(1997).			
[4]	SEQUENCE FROM N.A.			
RP	STRAIN=168;			
RC	MEDLINE=98044033; PubMed=9384377;			
RX	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,			
RA	Entlart K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grand G.,			
RA	Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinou S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Puigja C., Purnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiuchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,			
RA	Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			
RA	Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,			
RA	Winters P., Wipat A., Yamamoto H., Yanane K., Yasumoto K., Yata K.,			
RA	Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;			
RT	"the complete genome sequence of the Gram-positive bacterium Bacillus			
RL	subtilis. "			
RL	Nature 390:249-256(1997).			
[4]	SEQUENCE OF 15.			
RP	STRAIN=168 / IS58;			
RX	MEDLINE=9744398; PubMed=9298659;			
RA	Artelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,			
RA	Hecker M.;			
RT	"First steps from a two-dimensional protein index towards a response-			
RL	regulation map for Bacillus subtilis. "			
RL	Electrophoresis 18:1451-1463(1997).			
CC	-I- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.			

CC -!- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC
CC -!- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announcement/)
CC or send an email to license@isb-sib.ch).

CC	EMBL; AP001517; BAB06877.1; -.
CC	HSP; Q27743; LCEO.
DR	InterPro; IPR001557; L_LDH.
DR	InterPro; IPR001232; Mdh.
DR	InterPro; IPR00205; NAD_binding.
DR	InterPro; IPR001236; ldh.
DR	Pfam; PF00056; ldh; 1.
DR	Pfam; PF02866; ldh_C; 1.
DR	PROSITE; PRO0086; LLDHGRNASE.
DR	PROSITE; PS00068; MDH; FALSE_NEG.
KW	Oxidoreductase; Tricarboxylic acid cycle; NAD; Complete proteome.
FT	ACT_SITE 153 153
FT	BINDING 156 156
FT	ACT_SITE 180 180
FT	SEQUENCE 314 AA; 33680 MW; 02D132F11E3B8E34 CRC64;
CC	-----
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce or send an email to license@isb-sib.ch).

Query Match	5.3%	Score 100.5;	DB 1;	Length 314;
Best Local Similarity	22.6%;	Pred. No. 1.3;		
Matches 49;	Conservative 3;	Mismatches 74;	Indels 61;	Gaps 9;

```

QY 10 LQYAEIPKSEIPK-----VIEKAYIPVETLIIKEIPFGL-NTGYTLKFLPKDIID 60
Db   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 130 MTYVTYKSGFPKRNVIQSGVLDLTAR---FRFVAQELNLSVEDITGFVL----- 177
Db   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 61 LVKGGIASDLIETIGTSYTHAI--LPLLPSRVEAOVORDREKVELFVSPKG---FWLP 116
Db   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 178 ---GGHGDMDVPLIRYSYAGGIPLEKLLPOERIDAIVERTKGGGEIVGLLNGSAYAP 234
Db   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 117 ELAYDPPIPAILLKDNQYELFDAGEAMLFSAHLNSAIKPIKPLYPHLIKRAQEKRFY-- 174
Db   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 235 AASLAEMVEAILKDK-----KRVLPITIAVL-----EGEGYCYD 267
Db   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 175 ----TSYLLGLRELKAIKLVPEGKVTILKAVKDIEAV 207
Db   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 268 IYVGVPITILGGGIEKIVIELDTDEKATFAKSIEV 304
Db   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

```

RESULT 10
NOEK_RHISN
ID NOEK_RHISN STANDARD; PRT; 474 AA.
AC P55356;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
01-OCT-2001 (Rel. 40, Last annotation update)
16-OCT-2001 (Rel. 40, Last annotation update)
Phosphomannomutase (EC 5.4.2.8) (PMM).
DE NOEK OR Y4AI.
OS Rhizobium sp. (strain NGR234).
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OC NCBI_TaxID=394;
XX [1]
SEQUENCE FROM N.A.
MEDLINE=97305956; PubMed=9163424;
Freiburg C.A., Fellay R., Bairoch A., Rosenthal A.,
Perret X.
"RA Molecular basis of symbiosis between Rhizobium and legumes.";
"RA Nature 387:394-401(1997).
"-! -! CATALYTIC ACTIVITY: D-mannose 1-phosphate = D-mannose 6-phosphate.
"-! -! PATHWAY: BIOSYNTHESIS OF THE FUCOSE MOIETY OF THE NOD FACTOR.
"-! -! SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY
CC CC

```

```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AE000064; AAB91606.1;
DR InterPro: IPR001485; PG_PMM_mutase.
DR Pfam: PF00408; PGM_PMM_1;
DR Pfam: PF02878; PGM_PMM_1;
DR Pfam: PF02879; PGM_PMM_1;
DR Pfam: PF02880; PGM_PMM_1;
DR PROSITE: PS00710; PGM_PMM; 1.
KW Isomerase; Phosphorylation; Nodulation; Plasmid.
FT ACT_SITE 102 102
FT SEQUENCE 474 AA: 49670 MW: 6965310CEA96A22B CRC64;
SQ
Query Match 5.3%; Score 99; DB 1; Length 474;
Best Local Similarity 20.3%; Pred. No. 2.9;
Matches 80; Conservative 59; Mismatches 128; Indels 128; Gaps 16;
QY 15 IPKSETPKVEKA-----YIPVIETLIKEIPF 42
DB 114 LPDGEINKADEQAITAEQLSADATRVCEGRGADHSSEAVDFYIQRYETLLPKSLGK 173
QY 43 GLNITGYTLKFLPKDIDLVKGGIADSLIEITGTSYTHAILPLPLSRVAVQVORDREYK 102
DB 174 GLKIGLYQHSSVARDILTYLEHGANNV-----PVGRSEVFIPVDTEA- 217
QY 103 EELFEVSPK--GFWLPYELADPIIPA-----ILKNDGYEYLFAD-----GEAML 144
DB 218 --ISAATCKMLAAWAKAFADAFDAIVSSDADADRLPLDRTGTPLRGDGLLGLICARLEAKL 275
QY 145 FSAHL--NSAIKPKPL-----YPHLKAKQREKRFYISYLLGLRELKRAIKLVFE- 193
DB 276 IATPITSNGIEAASGVVYRVTRVSGPYVIAAMTEAVAKQKQVWGF--EANGVYMLGSNF 334
QY 194 --GKVTLKAVKDIEAV--PVWVAVNTAV-----MLGIRLPLMNPVKVASHWIEDKNI 242
DB 335 SFGGASLPALPTRDCLVPLIAALHMAVEAKTPLSGIVAMHRLPVALSGRIENYFPDRSDA 394
QY 243 L-----LYGTDEFTGYRDIAGYRMSVGLLEVIDELNSLCLPSELKHSR 289
DB 395 LVAFKASKANVSHLSRIGRVAGTDDVDGLRLTFEG-----GR 433
QY 290 ELYLRTSSWAPDKSLRWRE--DEGNARLNMLSYNM 323
DB 434 ILHIRPSGNAPE--LRCYVEADDPDAAEHLAQL 466
RESULT 11
RNR_AQUAE
ID RNR_AQUAE STANDARD; PRT; 705 AA.
AC 067834;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonuclease R (EC 3.1.-.-) (RNase R) (VacB protein homolog).
GN RNR OR VACB OR AQ_2046.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aufay M., Huber R.,

```

```

RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358 (1998).
CC -!- FUNCTION: 3'-5' EXORIBONUCLEASE THAT PARTICIPATES IN AN ESSENTIAL
CC CELL FUNCTION. ACTS NONSPECIFICALLY ON POLY(A), POLY(U) AND
CC RIBOSOMAL RNAs (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SI MOTIF DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AE000769; AAC07792.1;
DR InterPro: IPR004476; 3_prime_RNase.
DR InterPro: IPR002059; Cold_shock.
DR InterPro: IPR001900; Ribonuclease_II.
DR InterPro: IPR003029; SI.
DR Pfam: PF00575; SI; 1.
DR Pfam: PF00773; RNB; 1.
DR SMART: SM00357; CSP; 1.
DR SMART: SM00316; SI; 2.
DR TIGRFAMs: TIGR00358; 3_prime_RNase; 1.
DR PROSITE: PS01175; RIBONUCLEASE_II; 1.
DR PROSITE: PS01267; SI; 1.
KW Hydrolase; Exonuclease; Nuclease; RNA-binding; Complete proteome.
FT DOMAIN 615 696
FT SI MOTIF.
FT SEQUENCE 705 AA: 81537 MW: 2C08587229D1B05E CRC64;
SQ
Query Match 5.2%; Score 98.5; DB 1; Length 705;
Best Local Similarity 16.3%; Pred. No. 5.4; Indels 195; Gaps 22;
Matches 76; Conservative 80; Mismatches 80;
QY 5 VFHGNLQVAEIP---KSE--IPKVIKAYIPVIETLIKE-----IPGLNITGYTLKF 53
DB 100 VFHGDVVRKAKVTEFGKKEVRILKLRKAKDIAKVVVFEDQCYVWPLDEN-AHHRILL 158
QY 54 LPKDIIDLKVGIA-----SDLEIIG-----TSVT 79
DB 159 SKKDCQKLKEGVEVVLKITQFTPKSPARGKVIEVGNPKKFAIDVIRKYNLPTSP 218
QY 80 HAIL---PLLPLSRVAVQVORDREYKEL-----FEVSPKGFV-----114
DB 219 EKVIKEVEAIPPEIEPEEKRRDLREQLCFTIDPEKAGDFDDAVAIELTPGYKLYVH 278
QY 115 LPYELADPIIPAALKD-----NGEYLFADGEAMLFSAHLNSAIKPKLYPHLI 164
DB 279 IADVS-----YVREGTETDEKAYKRGFTYFPDRALHMLPEKLSAKLCSLRP-----326
QY 165 KAOREKFRFYISYLLGLRELKRAIKLVFEGKVTAKVKDIEAV---PVWVAVNTAVMLGI 221
DB 327 ---NEDKLAF-----TVWVFDESGLNKKAYDIYESVIRSKARLTNEALALIV 371
QY 222 G-----RLP-LMNPKKVAS-----W-----235
DB 372 GDPALKEKFPNLVEPLRMETLYRILSRKRWEMSGIDFDLPEAEVIVDEYGEPTAIYPYE 431
QY 236 -----IEDKNILLYCTDIEFTGYRDIAGYR-----SVGELLEVIDELNS 276
DB 432 RHVAHRIIEHFMTISANETVALH---LEHAGYPL--YRVHEPPDEEKVENLLEIGLGY 486
QY 277 ELCLPSELKHSGRLEYLRTSSWAPDKSLRWREDEGNARLNMLSY 321
DB 487 KVKRPH-----YTPKFFQKLIIEDEGRPEENLVRF 517
RESULT 12
LONZ_BORBU

```

```

ID DB 682 -LAYSIVKTYSSKLNFDVKESPEIHLHPGEGATPKDPSAGITTATATASILSDKKVPDLD 740
AC 051558;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent protease La homolog (EC 3.4.21.-).
GN B80613.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001162; AAC66962.1;
CC MEROPS; S16; UPK;
CC TIGR; B80613;
CC InterPro; IPR003593; AAA_Atpase.
CC InterPro; IPR003959; AAA_Atpase_cent.
CC InterPro; IPR003111; LON.
CC InterPro; IPR001984; Lon_endopep.
CC InterPro; IPR004815; Lon_fam.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02190; LON; 1.
CC PRINTS; PR00830; ENDOLAPTASE.
CC SMART; SM00382; AAA; 1.
CC SMART; SM00464; LON; 1.
CC TIGRFAMs; TIGR00763; lon; 1.
CC PROSITE; PS01046; LON_SER; 1.
KW Hydrolase; Serine protease; ATP-binding; Complete proteome.
FT NP_BIND 369 376
FT ACT_SITE 719 719
FT ACT_SITE 719 719 BY SIMILARITY.
SQ SEQUENCE 813 AA; 92312 MW; 5129AA1498C5D0F6 CRC64;

Query Match 5.2%; Score 98.5; DB 1; Length 813;
Best Local Similarity 21.7%; Pred. No. 6.6; Indels 95; Gaps 18;
Matches 72; Conservative 52; Mismatches 113;

QY 12 YAEIPKSE-----IPKVIKAYIPVIEETLKEEIPFGNITGYT----- 50
DB 507 YSVIEKLEIAKIFLIPSIIKESFLSKVIRIEDDVIFNL-IRNYTMESVGLKRVLTNL 565
QY 51 LKFLPKDII-----DLVKGGI--ASDLIEIIGTSYTH-----AILPLLSRVVQVOR 97
DB 566 IRLVRELLIYVSKDQIIKGNFYSPSSLIHGNLSLTHDPDIPGIYKLIININNYVY-- 623
QY 98 DREVKEELFEVSPKGF-----W-----LP-----ELAYDPTIIPAILKDNQY 135
DB 624 DTEEDNLIDKIDSSGFVGLATWNYGGTVLPVEATKFEKKDIIITSLGAIMKESQA-- 681
QY 136 LPADCEAMLFSAHLNSAIKPKIPKIPYPLIKAKQREKFRFYISYLLG-----LRELKAIK 189

Query Match 5.2%; Score 97.5; DB 1; Length 312;
Best Local Similarity 21.8%; Pred. No. 2.1;
Matches 47; Conservative 34; Mismatches 76; Indels 59; Gaps 9;

QY 10 LOVAETPKSEIPK-----VIEKAYIPVIEETLKEEIPFGNITGYTILKFLPKDII 61
DB 130 MSTYTFKESGFPKRNVRIGQSGVLDTAR---FRTFVAQE-----LNIS-----VKDVTGF 175
QY 62 VKGGLASDLIEIIGTSYTHAI--LPILPLSRVQVQVORREVKKEELFEVSPKGF--PWLP 117
DB 176 VLGGHGDMDWPIVRYSVAGIPEKLEKIPKDRDLAIKVERKGGGEIVNLGNSAIYAPA 235
QY 118 LAYDPIIPAILKDNQYEYLFADGEAMLFSAHLNSAIKPKIPKIPYPLIKAKQREKFRFY-- 174
DB 236 ASLAEMVEAIVKQD-----RRILPAITYL-----EGEYGEYEG 268

RESULT 13
MDH_BACTC
AC Q9X4K8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN MDH.
OS Bacillus thermodenitrificans.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=33940;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=00462;
RA Williams R.A.D., Welch S.G., Alawadhi S.A.;
RT "Properties and primary structure of a thermostable L-malate
RT dehydrogenase from 'Bacillus thermodenitrificans'.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -1- SIMILARITY: BELONGS TO THE LDH FAMILY MDH SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF114423; AAD28555.1;
CC HSSP; Q27743; ICET.
CC InterPro; IPR001557; L_LDH.
CC InterPro; IPR001252; Mdh.
CC InterPro; IPR00205; NAD_binding.
CC InterPro; IPR001236; ldh.
CC Pfam; PF00056; ldh; 1.
CC Pfam; PF02866; ldh; 1.
CC PRINTS; PR00086; LIDHGRGNASE.
CC PROSITE; PS00068; MDH; FALSE_NEG.
KW Oxidoreductase; Tricarboxylic acid cycle; NAD.
FT ACT_SITE 153 153
FT BINDING 156 156
FT ACT_SITE 180 180
FT ACT_SITE 180 180 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
FT ACT_SITE 180 180 PROTON-RELAY (BY SIMILARITY).
SQ SEQUENCE 312 AA; 33682 MW; 68C691C462EFF452 CRC64;

Query Match 5.2%; Score 97.5; DB 1; Length 312;
Best Local Similarity 21.8%; Pred. No. 2.1;
Matches 47; Conservative 34; Mismatches 76; Indels 59; Gaps 9;

QY 10 LOVAETPKSEIPK-----VIEKAYIPVIEETLKEEIPFGNITGYTILKFLPKDII 61
DB 130 MSTYTFKESGFPKRNVRIGQSGVLDTAR---FRTFVAQE-----LNIS-----VKDVTGF 175
QY 62 VKGGLASDLIEIIGTSYTHAI--LPILPLSRVQVQVORREVKKEELFEVSPKGF--PWLP 117
DB 176 VLGGHGDMDWPIVRYSVAGIPEKLEKIPKDRDLAIKVERKGGGEIVNLGNSAIYAPA 235
QY 118 LAYDPIIPAILKDNQYEYLFADGEAMLFSAHLNSAIKPKIPKIPYPLIKAKQREKFRFY-- 174
DB 236 ASLAEMVEAIVKQD-----RRILPAITYL-----EGEYGEYEG 268
```



```
QY 175 ---ISVLLGLRELKATKLVFEKVTLLKAVKDEAV 207
Db 269 YLGVPTILGGNGIEKVIETLETEDEKAALAKLSLEV 304

RESULT 14
UDPG_PVRPY
ID CARB_HELPY STANDARD; PRT; 471 AA.
AC O64459;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) (UDP-glucose
DE pyrophosphorylase) (UDPGP) (UGPase).
OS Pyrus pyrifolia (Japanese pear) (Pyrus serotina).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Rosales; Rosaceae; Maloideae; Pyrus.
ON NCBI_TaxID=3767;
RP SEQUENCE FROM N.A.
RA Kiyozumi D., Ishimizu T., Nakanishi T., Sakiyama F., Morioka S.;
RA "Molecular cloning and nucleotide sequencing of a cDNA encoding UDP-
RT glucose pyrophosphorylase of Japanese pear (Pyrus pyrifolia Nakai).";
RL (In) Plant Gene Register PGR99-006.
CC -!- FUNCTION: PLAYS A CENTRAL ROLE AS A GLUCOSYL DONOR IN CELLULAR
CC METABOLIC PATHWAYS.
CC -!- CATALYTIC ACTIVITY: UTP + alpha-D-glucose 1-phosphate =
CC diphosphate + UDP-glucose.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC UDGP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL; AB013353; BAA25917.1;
CC InterPro; IPR002618; UDGP.
DR Pfam; PF01704; UDGP.1.
KW Transferase; Kinase; Nucleotidyltransferase.
SQ SEQUENCE 471 AA; 51845 MW; CE5523CE35E13B40 CRC64;

Query Match 5.1%; Score 96.5; DB 1; Length 471;
Best Local Similarity 23.3%; Pred. No. 4.4;
Matches 65; Conservative 46; Mismatches 87; Indels 81; Gaps 17;

QY 16 PKSEIPKVIKAYIPVETLLKEEIPGLNIT-GYTLKFLPKDLIDLVKGIASDL---- 70
Db 67 PTPDEPEIKK-----LDLKLVLKNGLGITMGCT---GPKSVIEVRNGLTFLDLIVIQ 119
QY 71 TEICTSTYTHAILPLLP-----SRVEA-----QVQDRVKEELFEV 108
Db 120 TENLNKKGSCV-PLLLMNSFNTHDDTKIVKYSKSNVQHTFNQSQYPLRVDFESPL 178
QY 109 SPKG-----FWPELAYDPIIPAILKONGEYLFADGEAMLFSAHLNS--AIKPIKPLYP 161
Db 179 PSKQGTQKDGWYPP-GHGDVFPSLKNSCKGLDLLLSQGREYVFVIANSDNLGAVVDKILH- 236
QY 162 HLIKAQREKRRFYSYLL-----GLRELKAIKLVFEKVTLLKAV-----KDIE 205
Db 237 HLIQKNE-----YCMVEVTPKTLADYKGTGLISYEGRVQLLEIAQVDPQHVNEFKSIE 289
QY 206 AVPV-----WVAVNT-----AVMLIGRLPLMNPVKV 232
Db 290 KFKIFNTNLWNLNAIKRLVEADALKWEIIP--NPKEV 326

RESULT 15
ID CARB_HELPY STANDARD; PRT; 1085 AA.
AC O25577;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR HP0919.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RP SEQUENCE FROM N.A.
RA STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds three manganese ions (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL; AE000601; AAD07963.1;
CC HSSP; P00968; 1A9X.
DR TIGR; HP0919;
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF00289; CPase_L_chain; 2.
DR Pfam; PF02786; CPase_L_D2; 3.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR0098; CPASE.
DR PROSITE; PS00866; CPASE_1; FALSE_NEG.
DR PROSITE; PS00867; CPASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 399 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 400 552 OLIGOMERIZATION DOMAIN.
FT DOMAIN 553 951 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 952 1085 ALLOSTERIC DOMAIN.
FT REPEAT 1 552
FT REPEAT 553 1085
FT NP_BIND 151 208 ATP (POTENTIAL).
FT NP_BIND 301 350 ATP (POTENTIAL).
```

```
FT METAL 283 283 MANGANESE 1 (BY SIMILARITY).
FT METAL 297 297 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 299 299 MANGANESE 2 (BY SIMILARITY).
FT METAL 830 830 MANGANESE 3 (BY SIMILARITY).
FT METAL 842 842 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1085 AA; 120073 MW; 696E124A50DB34C1 CRC64;

Query Match 5.1%; Score 96.5; DB 1; Length 1085;
Best Local Similarity 20.0%; Pred. No. 14;
Matches 80; Conservative 61; Mismatches 124; Indels 135; Gaps 18;

QY 16 PKSEIPKVIKAYIPVIEITLKEEIPFGNITGYTLKFLPKDIDLVKGGIASDLIEIG 75
Db 689 PKNGMAKSVDEAYSA-----NVIGFPIIVRPSYVL-----GGQHMQILENIE 731
QY 76 -----TSYTHAI-----LPLPLSRVEAQVQDREV---KEELF-----EVSPKGF-- 113
Db 732 ELRHYLESVTHALEISPKNPILLIDKFEKAVELVDVAICDKKEVYIAGILQHIEAGIHS 791
QY 114 -----WPELAYDPTIPAILKONGYEYLFADGEAML---FSAHLNSAIKPIKPLYPHLIK 165
Db 792 GDSACFIPSTLSPEILDEIERVSAKIALHLGVVGLLNQFAVHONS-----LYLIEVN 844
QY 166 AQREKFRFVSYLLG-----LRELKAIKLVFEGKVTLKAVKDI--EAVPVVW 211
Db 845 PRASRTVPFLSKALGVPLAKVATRVWVLEDLREALKF-YDKKNIVGYSGVYKPKMPHFV 903
QY 212 AVNTAV-----MLGIGRL-----PLMNPKKVASW 235
Db 904 ALKEAVFPFNKLYGSDILGPEMKSTGEVMGIARSLGLAFFKAQACFPNIKNGLI FVS 963
QY 236 IEDKDNILLYGTDIEFIGYRDIAGYRMSVEGLEVIDELNSELCLPSELKHSGRELYLRT 295
Db 964 IKDKD-----KEACVLMKRLVOLGPELC-ATEGTHKALEK----- 998
QY 296 SSWAPDKSLRIWREDEGNARLNMLSYNMRGELAFIAENS 335
Db 999 ---AGVKSLKLVKISEG--RPNIMDLMMNGEISMAINTSD 1033
```

Search completed: November 18, 2002, 07:25:29
Job time : 18 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 07:24:24 ; Search time 22 Seconds
(without alignments)
1590.588 Million cell updates/sec

Title: US-09-619-032-4
 Perfect score: 1880
 Sequence: 1 MRALVFHGNLOYAEIPKSEIRRLDAAFRATYNDWRGENGP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database : PIR_73:**
1: pir1:**
2: pir2:**
3: pir3:**
4: pir4:**
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1506	80.1	364	2	E71144	hypothetical prote	
2	167	8.9	529	2	AC2112	hypothetical prote	
3	161.5	8.6	529	2	S76831	hypothetical prote	
4	161.5	8.6	633	2	G71241	probable alpha-amylase A (EC 3.2.2.)	
5	159.5	8.5	686	1	ALDYAT	uncharacterized co	
6	151	8.0	527	2	F97197	alpha-amylase (EC	
7	146.5	7.8	467	2	B64501	alpha-amylase (or	
8	141.5	7.5	655	2	E75206	alpha-amylase (EC	
9	140.5	7.5	649	2	A49512	hypothetical prote	
10	129.5	6.9	560	2	D71011	hypothetical prote	
11	127	6.8	923	2	G38826	hypothetical prote	
12	126	6.7	1362	2	A75207	anoylpullulanase p	
13	123.5	6.6	744	2	AC1970	hypothetical prote	
14	119	6.3	602	2	C75120	hypothetical prote	
15	115.5	6.1	324	1	B69553	methanol dehydroge	
16	111	5.9	526	2	D71334	conserved hypoteth	
17	109.5	5.8	447	2	E90250	alpha amylase [imp	
18	109	5.8	312	2	I40383	malate dehydrogena	
19	109	5.8	1069	2	AF1930	hypothetical prote	
20	107.5	5.7	902	2	E90270	conserved hypoteth	
21	104	5.5	526	2	B70859	hypothetical prote	
22	104	5.5	546	2	A12045	serine/threonine k	
23	103	5.5	704	2	G96587	hypothetical prote	
24	103	5.5	1150	2	G89881	pyruvate carboxyla	
25	101.5	5.4	684	2	T47694	probable serine/th	
26	101	5.4	4845	2	T31067	BIR repeat contain	
27	100.5	5.3	266	2	G36943	probable xylanase/	
28	100.5	5.3	314	2	F84044	malate dehydrogena	
29	100.5	5.3	619	2	D71361	probable alpha-amylase	

30	99.5	5.3	1181	2	T01799	hypothetical prote
31	99	5.3	503	2	T19319	hypothetical prote
32	99	5.3	636	2	S70877	hypothetical prote
33	99	5.3	815	2	G92766	mannose-1-phosphat
34	98.5	5.2	705	2	F70475	vacB protein (ribo
35	98.5	5.2	725	2	T44992	translation elonga
36	98.5	5.2	813	1	D70176	endopeptidase La (
37	98	5.2	510	2	B70368	L-aspartate oxidas
38	97.5	5.2	588	2	A61926	two-component resp
39	97	5.2	764	1	H70414	conserved hypothet
40	96.5	5.1	627	2	D84531	probable salt-indu
41	96.5	5.1	1085	2	G64634	carbamoyl-phosphat
42	96	5.1	530	2	F82271	transglycosylase,
43	95.5	5.1	1134	2	D75014	hypothetical prote
44	95	5.1	394	2	F75006	hypothetical prote
45	94.5	5.0	453	2	D84018	two-component resp

ALIGNMENTS

RESULT 1

E71144
 C:Species: *Pyrococcus horikoshii*
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 08-Sep-2000
 C:Accession: E71144
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hailkawa, Y.; Hino, Y.; Yamamoto, S.; Seno, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguni, A.
 M: Nucleic Acids Res. 5, 557-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
 A:Reference number: A71000; MUID:98344137; PMID:9679194
 A:Accession: E71144
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-364 <RAW>
 A:Cross-references: NID:g3236129; PIDN:BAA29442.1; PID:d1030385; PID:g3236129
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C:Genetics:
 A:Gene: PH0368
 C:Superfamily: *Pyrococcus horikoshii* hypotheoretical protein PH0368

Query Match	80.1%	Score	1506;	DB 2;	Length	364;			
Best Local Similarity	76.6%;	Pred. No.	2e-106;						
Matches	278;	Conservative	47;	Mismatches	38;	Indels	0;	Gaps	0;

QY	1	MRALVPHGNLOYAEIPKSEIPKVIKAYIPVITLKEEIPFGNITGYTLKFLPKDID	60
DB	1	MRALIFHGNLOYAEIPKHEISKVIEKSYFPTISLILKREIPFGNITGYSLQLPQELH	60
QY	61	LVKGGIASDLIEITIGTSYTHAILPLPLPSVEAQVORDREVKEELFEVSPKGFWLP	120
DB	61	LIKEGIESELEIIGTSYTHAILPLPLTLSRIEQAQIKDRDREIKEIFEIVSPGGFWLP	120
QY	121	DPITPAILKONGYEYIFADGEAMLFSAHLNSAIKIPKLYPHLITKAQERFRYISYLLG	180
DB	121	DPIIPAILRDNEYEYLFADGEAMLFNSHLNSAIKSIKPLYPLIKAQRGEGFVYLNLLG	180
QY	181	LRELRAIKULVFGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDK	240
DB	181	LRELKRAINLTFGGKVTLEAVKDIEAIPVWVSINIAITMLGAGRFPLMSPRKRVANWIKGD	240
QY	241	NILLYGTDIEFIGYRDITAGYRMSVEGLLEVIDELNSELCLPSLKHSGRELYLRTSSWAP	300
DB	241	EILLYGTDIEFLGYRSTAGHKAISGLMEVNLGGLGCLPRDIRHNGRRLYLRRTSSWAP	300
QY	301	DKSLRWREDEGNARLNMLSYNMRGELAFALAEANSDARGWEPLPERRLLDAFRAIYNDRWGE	360
DB	301	DKSLRWKEDEGNARLNMLTYCHMDGEFAFLAEANSDARGWEPLPERRLLDAFKAIFYWYRNE	360
QY	361	NGE	363
		II:	

Db 361 NGK 363

RESULT 2

AC2112

hypothetical protein alr2450 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AC2112

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

baena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AC2112

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-529 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA074149.1; PID:gl7131542; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr2450

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match

Best Local Similarity 8.9%; Score 167; DB 2; Length 529;

Matches 104; Conservative 55; Mismatches 162; Indels 174; Gaps 21;

Db 3 ALVFNHGNLOVAEIPKSE-----IPKVEKAYIP---VIETLKEEIPFGL-----44

Db 7 ALVLAHLFPVRHPESDYVLEEWLYEAITETIPLKVFGLKRDGIDFKITMSMTPL 66

Db 45 -----NITGYTLKFLPK-----56

Db 67 VSMLRDPLLOQRYEAHLAQLLELIELEGNAQNGHLRYLAHYATEFNEARQMRYNG 126

Db 57 DIIDLKVGGIASDLIEIGTSYTHAILPLPL--SRVEAQVQDRREVKEELFEVSPKGF 114

Db 127 DLVTAFAKQFQDSNNLEIITCGATHGYPLMKMYPQAVWAIQVACEHYEETEGRPKGIW 186

Db 115 LPELAYDPIIPAILKDNKGYEYLFADGEAMLFSALHNSAIKPKPL-----159

Db 187 LPECAVYEGVGLERMLADAGRLYFLTDGHGILY-ARPRPGTYAPIFTETGVAAGRDHES 245

Db 160 -----YPLIKFAQR-----EKFRVYSYLLGLRELKRAIKLVFEGKVT-----197

Db 246 SQQVWSEVGYPGAAYREFYKDLGWEAEYIKPYIMPNGQKNTGKIYH-KITGRGLG 304

Db 198 --LKAVKDIKAVPVWVAVNTAVMLGIGRLPLMNPKKVASWIEDKDNIL--LVGT-----247

Db 305 LSDKALYD---PYWAKKAA-----EHAANFYNNRERQAEHLGYIMQRPPI 347

Db 248 -----DIEFIGYRDITAG-----YRMSVGLLEVIDELNSE-----LCLFSELKH 286

Db 348 IVSPYDAELFGHWWGPGWFDYLFKRSWYDQGTAVMTHLADYLRNEPTQQVCRPSQ---404

Db 287 SGRELRYLTSSWAPDKSLRIWREDEGNA-----RLNMLSYNMKGELAFIAENSADRGWEPL 342

Db 405 -----SSWGKGFHGYWIL-NETNAWIYPHLHKAERMT-EISTL-EPEDELGWRAL 452

Db 343 PERRLDAFRAIYNDW 357

Db 453 NQAARELLLAQSSDW 467

RESULT 3

S76831

hypothetical protein - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: S76831

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O.; K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys

s.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76831

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-529 <KAN>

A:Cross-references: EMBL:D90916; GB:AB001339; NID:gl653715; PIDN:BAAL8743.1; PID:gl65

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match

Best Local Similarity 8.6%; Score 161.5; DB 2; Length 529;

Matches 101; Conservative 47; Mismatches 140; Indels 183; Gaps 19;

Qy 3 ALVFNHGNLOVAEIPKSE-----IPKVEKAYTPVI---ETLKEEIPFGLNT-----47

Db 7 ALVLAHLFPVRHPESDYVLEEWLYEAITETIPLIHVFEGLKRDGVDFKITMSMTPL 66

Qy 48 -----GYTLKFLPK-----56

Db 67 VSMLRDPLLOQRYEAHLAQLLELIELEGNAQNGHLRYLAHYATEFNEARQMRYNG 126

Qy 57 DIIDLKVGGIASDLIEIGTSYTHAILPLPL--SRVEAQVQDRREVKEELFEVSPKGF 114

Db 127 DLVTAFAKQFQDSNNLEIITCGATHGYPLMKMYPQAVWAIQVACEHYEENFGSPKGIW 186

Qy 115 LPELAYDPIIPAILKDNKGYEYLFADGEAMLFS-----HLNS 151

Db 187 LPECAVYEGVGLERMLADAGRLYFLTDGHGILYARPRPGTYAPIFTETGVAAGRDHES 246

Qy 152 -----AIKPIKPLYPHLIK-AQREKFRFYISVLLGLRELKRAIKLVF-----EGKVT 197

Db 247 QQVWSSQVGYPCDPVYREFYKDLGWEAEYIKPYIMPNGQKNTGKIYHKTITSRDGLS 306

Qy 198 LKAVKDIKAVPVWVAVNTAV-----MIGI-GRPLMNPKKVAS-----WI 236

Db 307 EKAWYD---PYWAKKAAEHASFMYNROQQVGHLSGIMGRPLPVVSPYDAELFGHWY 362

Qy 237 E-----DKDNI-----LLYGTDIETI-----GYRDIAGY-----260

Db 363 EGPWFIDYLFKRSWFDQGTAVMTHLADYLRNPHQVCRPSQSSWGKGFHETWLNNTNA 422

Qy 261 -----RMSVGLLEVIDELNSELCPLSELKHSRELYL-RTSSWA 299

Db 423 WIYPHLHKAERMTLSHREAVDELEEK-----ALNQAARELLLAQSSDNA 468

RESULT 4

G71241

probable alpha-amylase - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000

C:Accession: G71241

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hatai, Y.; Yamamoto, S.; Se

M.; Onfuk, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A71000; MUID:98344137; PMID:9679194

A:Accession: G71241

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-633 <KAW>

A:Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29262.1; PID:g3256579

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH0193

C:Superfamily: Dictyoglomus thermophilum amylase A

Query Match 8.6%; Score 161.5; DB 2; Length 633;
 Best Local Similarity 23.6%; Pred. No. 0.00026;
 Matches 96; Conservative 65; Mismatches 134; Indels 111; Gaps 26;

QY 8 GNQVAAEIPKSPKVIKAYIPVETLKEIP---FGLNITGYTLKFLPK---DIIDL 61
 DB 17 GNFEWI-----IKRAYEKAYRFFLET--EYPNMKVAVHISGVLVWELERNPEYIDL 68
 QY 62 VKGGIASDLIEITGTSYTHAILPLP-LSRVEAQVORDREVKEEL-FVSPKGFWMPELA 119
 DB 69 LKSLIKKQVELVAGFYEPILVAIPEDRVE-QIKLSKGMARKMGYEA--RGLWLTERV 125
 QY 120 YDPIIPAILKDNQYGYLFDAGEAMLSAHLNSA-IKPIKPLYPHLIKAKREKFRY---- 174
 DB 126 WEPVLKTLREAGIEVILD-----DYHFMASAGLSKEELFWPYTENGGEAIVFPDIDE 179
 QY 175 -ISVLLGLRLRKAIKLVF-----EGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLM 227
 DB 180 KLYLIPFRPVNETLEVHLSLADEDESKVAV-FHDDGKFGAWPGTHELVY----- 229
 QY 228 NPKKVASWI-----EDKNILLYGTDIEFIG-----YRDIAGY-RMSVEGL-- 267
 DB 230 -----ERGWLKEFFDRISDDKINMLYS---EYLSKFRPKGLVLYLPFIASFFEMSEWSLPA 282
 QY 268 -----LEVIDELNSLCLPSLKHSGRELYLRTSSWAPDKSLRIWREDEGN---ARLNM 318
 DB 283 RQAKLFFETTKKL-KELNLFKRY-----IFVRGGTW---KNP-LKYYPEGNYMHRMLM 332
 QY 319 LSYNMRGELAFLAENSARGWEPLPERRLDRAFIYND--WRGENG 362
 DB 333 LSKLLRNN-----PTARIPVLRAQCNDAYWHGVFG 362

RESULT 5
 ALDYAT
 amylase A (EC 3.2.1.-) - Dictyoglomus thermophilum
 C:Species: Dictyoglomus thermophilum
 C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 18-Jun-1999
 R:Fukushima, S.; Kamizono, A.; Horinouchi, S.; Beppu, T.
 Eur. J. Biochem. 174, 15-21, 1988
 A:Title: Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic
 A:Reference number: S00628; MUID:88225097; PMID:2453362
 A:Accession: S00628
 A:Molecule type: DNA
 A:Residues: 1-686 <FUK>
 A:Cross-references: EMBL:X07896; NID:g2688; PIDN:CAA30735.1; PID:g2689
 A:Accession: A34969
 A:Molecule type: protein
 A:Residues: 2-13 <FUK2>
 C:Genetics:
 A:Gene: amya
 C:Superfamily: Dictyoglomus thermophilum amylase A
 C:Keywords: glycosidase; hydrolase
 F:2-686/Product: amylase A #status experimental <MAT>

Query Match 8.5%; Score 159.5; DB 1; Length 686;
 Best Local Similarity 21.0%; Pred. No. 0.00041;
 Matches 85; Conservative 64; Mismatches 149; Indels 107; Gaps 19;

QY 8 GNQVAAEIPKSPKVIKAYIPVETLKEIP-EIIPFGLNITGYTLKFLPK---DIIDL 63
 DB 18 GNFDV-----IERAYEMSKYPLINFFKHPDPIINVHFSGLLLMLEKNHPEYFEK 71
 QY 64 GGIASDLIEITGTSYTHAILPLP-LSRVEAQVORDREVKEEL-FVSPKGFWMPELA 123
 DB 72 IMAERGQIEFVSGGFEYEPILPIIDKQVQIKLNKYIYDFQGTQPKGMWLAERVWEPH 131
 QY 124 IPAILKDNQYGYLFDAGEAMLSAHLNSA-----AIKPIKPLYPHLIK-A 166
 DB 132 LVKYTAEGIEVVDV-DAHFSVGLKKEEDLFYGLMEEQYKLAFFISMKLYLPFA 190
 QY 167 QREKFRYISVLLGLRLRKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPL 226

DB 191 DPEETIYLDKFAS--EDKSKIALFD-----DGEKFLWPDYRTVY----- 231
 QY 227 MNPKKVASWIED-----KDNILL-----YCTDIEFIG--YRDIAGYRMSVGLL--- 268
 DB 232 -----EGWLETFVSKIKENFLLVTPVNLVYTMQRVQPKGRIYLPFTASYREMMEWVLPPE 286
 QY 269 -----EVIDELNSLCLPSLKHSGRELYLRTSSWAPDKSLRIWRE-----DEGNARLN 317
 DB 287 AQKELELVEKLKTE-----NLW---DKFSYVKGFWRNFLAKYDESN----- 327
 QY 318 MLSYNMRGELAFLAENSARGWEPLPERRLDRAFIYND--WRG 359
 DB 328 ---HMOKMLLYVMKKVQDSPNEEVKEKAMEEVFGQANDAYWHG 368

RESULT 6
 F97197
 uncharacterized conserved protein, related to alpha-amylase/alpha-mannosidase CAC2414
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
 C:Accession: F97197
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: F97197
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-527 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK80369.1; PID:g15025430; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 C:Gene: CAC2414
 C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match 8.0%; Score 151; DB 2; Length 527;
 Best Local Similarity 18.9%; Pred. No. 0.0012;
 Matches 92; Conservative 63; Mismatches 119; Indels 212; Gaps 20;

QY 5 VFHGNLOYAEIPKSE-----IPKVIKAYIPVIB---TLIKEEIPFG----- 43
 DB 9 VLHSHMPFVRHPETRDSLEERWLFAMSECTYPLIEVYDNLKDKIKFRMTWTSITPPLMS 68
 QY 44 -----LNIITGYTLKFLPKDII-----DL 61
 DB 69 MLODEVLSRYLNLAKTIELSEKEILRTKNRREKNVALFYNNKRAENTLIKEYKYDNNL 128
 QY 62 VKGGIASD---LIEITGTSYTHAILPLP-LSR--VEAQVORDREVKEELFEVSPKGFWM 116
 DB 129 INAFKRYDRLGCVIELITCAATHALLPLILINROAVRAQIATGVQSYINTMGHEPNIGWLP 188
 QY 117 ELAYDPIIPAILKDNQYGYLFDAGEAMLSAHLNSA-IKPIKPLYPHLIKAKREKFRYIS 176
 DB 189 ECAYTYGIDNILSEFGIKYPISEGAIDVASP-----KPMY----- 224
 QY 177 YLLGLRELKRAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWI 236
 DB 225 -----GTNT-----PIAAPSQVCAFG 240
 QY 237 EDKNILLYGTDIEFIG-----YRDIAGYRMSVGLLEVIDELNSLCLPSLKH 286
 DB 241 RMDSD--SYQWSDFMGYPGDFNRYREFYRDI-GFELPMYIKPYINE--NGRIDTGFY 295
 QY 287 -----SGRE-LYLRTSS---WAPDKSLRIWREDEGN-ARLNM----- 318
 DB 296 YKITGNSGEKGIYNRENAMKKVWEHASHFASCRHQDQNAANNDKPPITTCYDTELYG 355
 QY 319 -----LSYNMRGELAFLAENSARGWEPLPERRLDRAFIYNDW 357
 DB 356 HWFEGPFDIFNAIRKSAEDWTSYELITPTTEYLKNNSMVQSSPSP-----SSW 404

QY 358 RGENGE 363
DB 405 -GENGD 409

RESULT 7

B64501
alpha-amylase (EC 3.2.1.1) - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
A:Accession: B64501
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinschenk, K.G.; Merrick, J.M.; Glodok, A.;
son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: B64501
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-467 <BUL>
A:Cross-references: GB:U67601; GB:L77117; NID:g2826439; PIDN:AAB99631.1; PID:g1592212; T
C:Genetics:
A:Map position: FOR1586396-1587799
C:Function:
A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A:Pathway: glycogen/starch degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 7.8%; Score 146.5; DB 2; Length 467;
Best Local Similarity 23.1%; Pred. No. 0.0023;
Matches 87; Conservative 66; Mismatches 129; Indels 95; Gaps 24;

QY 14 EIPKSEIPKVEKAYIPVETLIK--EIPFGLN--ITG-----YTLKFLPKD-IIDLKVG 64
DB 34 KLNKEVENKANKCYIPTNELLLEIDEDYFKVNSITGVFVEQALEF--NDYVLDLFDK 91
QY 65 GIASDLIEIGTSTHAILPLPL-SRVEAQVQDRVKEELFVSPKGLWELAYDPI 123
DB 92 LVKGNVELIAETVHHSLSLFTFEDIEDIEHMKMYKEIFGKAKVFRNELIYNNR 151
QY 124 IPAILKONGYEYLFADG-EAMLFSAHLNSAIKPKPLPHLIKAQREKFRYIS---YLL 179
DB 152 IAKIADLGFKAIFTEGIEKIL-----GWRSPNLYOSPGMKI 190
QY 180 GLRELKAIKLVFEGKVTYKAVKDIEAVP-----VWAVNTAVMLGIRLPLMNPVKV 232
DB 191 LLRNRLSDDIGFR-----FSARDWDQYPLTADKYAIWLASTPGEVINI-----YMDYETF 241
QY 233 AS--WIEDKDNILLYGTDI-EFIGYR--DTAGY-RMSVGLLEVIDELSELCLPSELKH 286
DB 242 GEHWKE-----TGIFEFLYLPFIEIAKHEHLEVNVNVEVDRLPR-----283
QY 287 SGRELXLR---TSSWA-PDKSLRIWREDEGNRLNMLSYNMRGELA-FLAENSADRGWEP 341
DB 284 --GEIYVHEFATISWADTERDVSALW--GN-KMQRISFEKULKDKCFIKENSKL-----333
QY 342 LPERLDAPRAIYNDWR 358
DB 334 ---KLNKFEDELYKMYK 347

RESULT 8

E75206
alpha-amylase (or 4-alpha-glucanotransferase) PAB0118 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
A:Accession: E75206
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001

A:Accession: E75206
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-655 <KAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49100.1; PID:g545
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: amyA; PAB0118
C:Superfamily: Dictyoglomus thermophilum amylase A

Query Match 7.5%; Score 141.5; DB 2; Length 655;
Best Local Similarity 21.9%; Pred. No. 0.0088;
Matches 88; Conservative 54; Mismatches 130; Indels 129; Gaps 21;

QY 25 EKAYIPVETLIKEIP---FGLNITGYTLKFLPK---DIIDLKVGGIASDLIEIGTGY 78
DB 28 EKAYRPFLEIL--EYYPNNMKVAIHISGILVEENKPKDYIDLKSLVRKGOVEIVVAGF 85
QY 79 THAILPLPLSRVEAQVQDRVKEELFVSPKGLWELAYDPIIPAILKONGYEYLF 138
DB 86 YEPVLAAPKEDRLEQIYLLKWKAKKI-GYDAKGLMLTERVQPELVKTLREAGIEYVV 144
QY 139 DGEAMLFSAHLNSA-IKPIKPLYPHLIKAQRE-----KRFY-----ISYLL 179
DB 145 D-----DYHFSAGLSKDOLFEPYITVFPIDEXLRYLIPRPPVDKVISYLH 198
QY 180 GLRELKAIKLVFEGKVTYKAVKDIEAVPVVAVNTAVMLGIRLPLMNPVKVSWIEDK 239
DB 199 STASEDESKVAVFH-----DDGEKFGIW-----PM-----TYEMVYK 231
QY 240 D-----NILLYGTDI-----FIGYDIAGY-----RMSVEG 266
DB 232 GWLREFFDRVSSDEAINIMLYSEYILQKPKGLVPLPIASYFEMSEWSILPAQAKLFV 290
QY 267 LLEVIDELNSELCLPSELKHSRELYLRTSSWAPKSLRIWREDCN---ARLNMLSYNN 323
DB 291 FVEKLELN-----MFEYRVFVGGIW---KNF-FYKYPEANTYMHKRLMLSLRL 337
QY 324 RGEALFLAENSADRGWELPERRLDAPRAIYND--WRGENG 362
DB 338 RDN-----PSARFVLAQCNDAWHGVFG 362

RESULT 9

A49512
alpha-amylase (EC 3.2.1.1) - Pyrococcus furiosus
C:Species: Pyrococcus furiosus
C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
A:Accession: A49512
R:Laderman, K.A.; Asada, K.; Uemori, T.; Mukai, H.; Taguchi, Y.; Kato, I.; Anfinsen,
J. Biol. Chem. 268, 24402-24407, 1993
A:Title: alpha-Amylase from the hyperthermophilic archaeobacterium Pyrococcus furiosus
A:Reference number: A49512; MUID:94043280; PMID:8226990

A:Accession: A49512
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-649 <LAD>
A:Cross-references: GB:L22346; NID:g347939; PIDN:AAA72035.1; PID:g347940
C:Genetics:
A:Gene: amyA
A:Start codon: GTG
C:Function:

A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A:Pathway: glycogen/starch degradation
C:Superfamily: Dictyoglomus thermophilum amylase A
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 7.5%; Score 140.5; DB 2; Length 649;
Best Local Similarity 22.8%; Pred. No. 0.01;
Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;

QY 25 EKAYIPVETLIKEIP---FGLNITGYTLKFLP---KDIIDLKVGGIASDLIEIGTGY 78
DB 28 EKAYRPFLEIL--EYYPNNMKVAIHISGILVEENKPKDYIDLKSLVRKGOVEIVVAGF 85

Db 29 EKCWPFLETL--EYPNMKVIAHTSGPLIEWLQDNRPEDIDLLRSIVKRGQVEIVVAGF 86
QY 79 THAILPLLSRVAQVORDREVE--ELFEVSPKGFWLPDLAYDPIIIPAILKDNQYEYL 136
Db 87 YEPVLASIP---KEDRIEQLRLKWEAKSGFGDARGVWLTFRVQWQPELVKTLKESGIDYV 143
QY 137 FADGEAMFSAHLNSAIKPIKPLY-PHLIKAORE-----KRFYVISYLLGLRELKRA 187
Db 144 IVD-----DYHFMASGLSEELIWPYTYDEGGGEVIAVFPIDELKLR---YLIPFRPVDKV 194
QY 188 IKL---VFEKGVTILKAV--KDIEAVPVWVAVTAVMLGIGRLPLMNPKKVASWI----- 236
Db 195 LEYLHSLIDGDESKVAVFHDDGKFGIWPCTYEWVY-----EKGLREFEEDR 241
QY 237 ---EDKNILLYGTDIE-----FIGYRDIAGYRMSVEGLLEVIDELNSELCLP----- 281
Db 242 ISSDEKINMLTYLEKYKPRGLVLPFIASY-----FEM-----SEWSLPAKQARLF 289
QY 282 ---SELKHSQ---RELJRTSSWAPDKSLRIWREDEGN---ARLNMLSYNMRGELAF 330
Db 290 VEFVNELVKGIKFEKYRVFVGGIW---KNF-FYKYPESNYMKRMLMVKSLVNN----- 341
QY 331 AENSARGWEPLPERLLDAFRAYND--WRG 359
Db 342 -----PEARKYLLRAQCNDAYWHG 360
RESULT 10
D71011
hypothetical protein PH1386 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: D71011
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, J.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: D71011
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-560 <KAW>
A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30492.1; PID:g3257809
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1386
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386
Query Match 6.9%; Score 129.5; DB 2; Length 560;
Best Local Similarity 25.6%; Pred. No. 0.057;
Matches 60; Conservative 28; Mismatches 81; Indels 65; Gaps 12;
QY 14 EIPKSEIPKVIKAYIPVETL-----IKKEIPFGLNITGYTLKPLPKDIIDL-----VK 63
Db 71 EYIKREFEYMERKLSMEEDLERFKDKLREAINF---MIGYF-----KDVSYNWKSID 122
QY 64 GGIASDL-----IEIGTSYTHAILPLPLSR-VEAQVORDREYKEELFEVSPKGF 114
Db 123 GNILGKFLQDEGVEVITSAATHGYLPLGRDEATEAOLLNGIKYKYFGKRPGRGIW 182
QY 115 LPELAYDPI-----IPAILKDNQYEYLFADGEAMLFSAHLNSAIKPIK 158
Db 183 LPFCAYRDPGLWKSPTGCEVKKRGIEHFLKKGIEFFVE-----SHLIDK-GPVSL 234
QY 159 LYPHLIKAOREKFRFYISYLLGLRELKRAIKLVFEGKVTLKAVKDIEAVPVWA 212
Db 235 RYGNILPA-KTKRSTLRPYF-----LKNGIADVARNRET-----GIOWSA 274
RESULT 11
G83826
hypothetical protein BH1415 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83826
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G83826
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-923 <STO>
A:Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA05134.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1415
Query Match 6.8%; Score 127; DB 2; Length 923;
Best Local Similarity 23.8%; Pred. No. 0.18;
Matches 54; Conservative 31; Mismatches 70; Indels 72; Gaps 11;
QY 3 ALVFHGNLOYAEPKSE-----IPKVEIAKIPVETILKEEIPFGLNI----- 46
Db 7 SLVLHAHLPYVRHQEEDRLERLWLFAMSETYIPLLWALEKLPLVKAHTVTSFTPPVMEML 66
QY 47 -----TGY-----TLKFLPKD-----IIDLVKG-----G 65
Db 67 SDPLVQTRYNLHLENTEQLLKKEKRTNDQRTONLVQFYKQRYEKLKATFLQWRNLLIG 126
QY 66 IASDLIE-----IIGTSYTHAILPLPLSR-VEAQVORDREYKEELFEVSPKGFWLPDLA 119
Db 127 FRS-LMENEQCTLMTSAATHAFPPYLTREATRAOVRHGFIACFEQHFQKGLGFWLPECA 185
QY 120 YDPIIPAILKDNQYEYLFADGEAMLF---SAHLNSAIKPIKPLY-PH 162
Db 186 FSGVDRIILFEFGIRYTFVDEHAULTADPTPHKGSS-----APIYSHP 228
RESULT 12
A75207
amylopullulanase PAB0122 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A75207
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: A75207
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1362 <KAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49104.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: apu; PAB0122
Query Match 6.7%; Score 126; DB 2; Length 1362;
Best Local Similarity 21.8%; Pred. No. 0.37;
Matches 83; Conservative 56; Mismatches 142; Indels 100; Gaps 16;
QY 32 IETLKEEIPFGLNITGYTLKPLPKDIIDLKVGGIASDLIEIIGTSYTHAILPLLP----- 87
Db 236 VETVLHQM-WLLNHT-----FEEHEKINLLG---NGNVEVTVPTHTPIGPIILNDFGW 286
QY 88 LSRVEAQVORDREYKEELF---EVSPKGFWLPDLAYDPIIIPAILKDNQYEYLFADG---E 141
Db 287 YEDFDAQVKKANELYKEYLGAGKVTPKGWAESALNDKLTLELAENGKWKWMTDQVLVE 346
QY 142 AMLFAHLNSAIKPIK-----IKPLYPHLIKA 166
Db 347 KLGVPKTIETSYYPWVAQFGDKKIYLFPRNHDLSDRVGFYAGMNOYDAVKNEVEELKI 406
QY 167 QREKRFYISYLLGLRELKRAIKLVFEGKVTLKAV-KDIEAVPVWVAVTAV-----ML 219

Db 407 OKONYDGLSVYVITLGEWNEWHPDGLFLEELYRQLEELQKGLIRIVTPSEYIEMF 466
QY 220 G--IGRLPLMNPKKVASWIEDKDNILLYGTDIEFYGRIAGYRMSVEGLLEVIDELANSE 277
Db 467 GDKANKLITPKMKKRLDTEDNVALLKAKTL-----GELYDMVGVT----- 509
QY 278 LCLPSELKISUGRELYLRTSSWAPDKSLRTWREDEGNARLNMLSYNNRGELAFLEANS-D 336
Db 510 -----EMQWPESSWI-DGTLSTW---IGEPQENIAWYLYLARKALFENKONV 553
QY 337 RGWEPLPERRLDFAIRAIYNDW 357
Db 554 KDNWKAYEY---LFRAGSDW 571
RESULT 13
AC1970
hypothetical protein alr1310 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC1970
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yanada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1970
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073267.1; PID:g17130657; GSPDB:GN000179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1310

Query Match 6.6%; Score 123.5; DB 2; Length 744;
Best Local Similarity 22.6%; Pred. No. 0.24;
Matches 79; Conservative 49; Mismatches 111; Indels 111; Gaps 19;
QY 68 SDLIEIGTSYTHAILPL-----LPLSR-----VEAQQVQREVKEELF 106
Db 208 SGQLEVTTPYTHPLPLDNTSGRVAVPNNALPESFQWSEIDPRLRKAWELYTERF 267
QY 107 FEVSPKGFWLPCLAYDP-IIPAILKONGVEYLFAD-----GEAMLFSAHLSAIPKIPK-L 159
Db 268 GQEPKGLPSEQSVSDILPYIIR-KGFQWICSDAVLGLWTLKHFHFRDAGNVQOPELL 326
QY 160 Y-PHLIKAQ-----REKFRYISYLLGLRELKAKLVEGKVTILKAVKDIEAVPVW 210
Db 327 YRPYRLATPAGDLAIVFRDHR---LSDLIGF-----TYGAMPKQAAADL 368
QY 211 VAVNTAVMLGIGRLPLMNPKKVASWIEDKDNILLYGTD-----IEFYGRIAGY-RMSVE 265
Db 369 VGHQAIAKMQRERFSEQPWLVTIALDGENCWFEFFQDGKPFLEAL-YQSLNNEPHIKLV 427
QY 266 GLELWIDELNSCLPSELKHSRELYLRTSSWAPDKSLRTWREDEGNARLNMLSYNNRG 325
Db 428 TVSEFIEEPATATIPAEQLHSG-----SWV-DGSFTTWIGDP----- 464
QY 326 ELAFLAENSARGWEPPLPERRL-----DAFRAIY-----NDW 357
Db 465 -----AKN---RWDYLTEARIMLANHPENTEENNPPEAWELIYAEGSDW 506

RESULT 14
C75120
hypothetical protein PAB1857 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: C75120
R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: C75120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-602 <KAW>
A:Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49676.1; PID:g545
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1857
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match 6.3%; Score 119; DB 2; Length 602;
Best Local Similarity 26.8%; Pred. No. 0.39;
Matches 42; Conservative 20; Mismatches 59; Indels 36; Gaps 4;
QY 14 EIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPK-----DII 59
Db 71 EYKAEFNRYMERKI-----KLMKEDLKADGKLRNAIEFMKYFEDVYEWKINGDII 125
QY 60 DLVKGGIASDLIEIGTSYTHAILPLPLSLR-VEAQVQREVKEELFEVSPKGFWLPCL 118
Db 126 GRFKQLQDEGFVEIITSAATHGYLPLLRDEADDAQLITGIRVEYKEFGKPKRGWLPCL 185
QY 119 AYDPI-----IPAILKDNIGVEYLFAD 139
Db 186 AYRPDGLWKSPTGEIKWRKGIEHFLKKGLEFFVE 222

RESULT 15
B69553
methanol dehydrogenase regulatory protein (moxr) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: B69553
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiaich, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: B69553
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-324 <KLE>
A:Cross-references: GB:AE001108; GB:AE000782; NID:g2689431; PIDN:AAB91247.1; PID:g265
C:Superfamily: methanol dehydrogenase regulatory protein

Query Match 6.1%; Score 115.5; DB 1; Length 324;
Best Local Similarity 21.1%; Pred. No. 0.3;
Matches 69; Conservative 62; Mismatches 121; Indels 75; Gaps 14;
QY 1 MRALVFHGNLOYAIEP---KSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKD 57
Db 35 LAAALINGNITLFDYPLGLKTLAKVFAVY-----ICADYRRVOFTPDLLPSD 82
QY 58 IIDLVKGGI---ASDLIEII-GTSYTHAIL-----PLPLSRVEAQVQREVKEEL 105
Db 83 II-----GVKWRGDRFEFVKGGIFITNVLDAEINRSPKTOAALLEAMEEKQITVEGET 137
QY 106 FEVSPKGFWLPCLAYDPIIPAILKDNIGVEYLFADGFAMLFSAHLSAIPKIPKLYPHLIK 165
Db 138 FSLSMPPFVL--ATQNPI-----EQEGTYPLPEAQMDRMLMRPGYPEESIE 182
QY 166 AQREKFRYISY-----LLGLRELKAKLVEGKVTILKAVKDIEAVPVW 211
Db 183 EEMEILRRISRKDDPTDEVPVSLFTFRQQDAVEAVYVDKSLIKYSELVRA---T 239
QY 212 AVNTAVMLGI---GRPLMNPKKVASWIEDKONIL---LYGTDIEFYGRIAGYRMSVE 265

Db 240 REHELVELGSSPRGGLALLKLARALAVMDGRDFVIPDDVKRVAVEALAHRVILKFEYAVE 299

QY 266 GLL--EVIDELNSLCLPSELKHSGRE 290

Db 300 GLRAEEVVEEILNSVRVP---KYEAGE 323

Search completed: November 18, 2002, 07:25:58
Job time : 27 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 07:26:04 : Search time 9 seconds

(without alignments)
609.129 Million cell updates/sec

Title: US-09-619-032-4

Perfect score: 1880

Sequence: 1 MRALVPHGNLQYAEIPKSEI.....RLDAFRAIYNDRWGENGEP 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications-AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB_PEP.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB_PEP.*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB_PEP.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB_PEP.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB_PEP.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB_PEP.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB_PEP.*
8: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB_PEP.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB_PEP.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB_PEP.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB_PEP.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB_PEP.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB_PEP.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1868	99.4	364	9	US-10-112-418-4 Sequence 4, Appli
2	1868	99.4	364	9	US-10-114-083-4 Sequence 4, Appli
3	1868	99.4	364	10	US-09-886-400-4 Sequence 4, Appli
4	1868	99.4	364	12	US-10-112-357-4 Sequence 4, Appli
5	1868	99.4	364	12	US-10-114-403-4 Sequence 4, Appli
6	1868	99.4	364	12	US-10-116-606-4 Sequence 4, Appli
7	1868	99.4	364	12	US-10-112-331-4 Sequence 4, Appli
8	1868	99.4	364	12	US-10-112-377-4 Sequence 4, Appli
9	1868	99.4	364	12	US-10-116-581-4 Sequence 4, Appli
10	1868	99.4	364	12	US-10-112-442-4 Sequence 4, Appli
11	103	5.5	1073	10	US-09-815-242-12361 Sequence 12361, A
12	103	5.5	1147	10	US-09-815-242-5468 Sequence 5468, Ap
13	91	4.8	813	10	US-09-732-180-7 Sequence 7, Appli
14	91	4.8	844	10	US-09-732-180-2 Sequence 2, Appli
15	89.5	4.8	1688	10	US-09-815-242-5654 Sequence 5654, Ap
16	89.5	4.8	2397	10	US-09-815-242-12265 Sequence 12265, A
17	88.5	4.7	764	10	US-09-828-466-4 Sequence 4, Appli
18	88.5	4.7	993	10	US-09-894-998-50 Sequence 50, Appli
19	88.5	4.7	1037	10	US-09-894-998-54 Sequence 54, Appli

ALIGNMENTS

RESULT 1

US-10-112-418-4

; Sequence 4, Application US/10112418

; Patent No. US20020155486A1

; GENERAL INFORMATION:

; APPLICANT: DIVERSA CORPORATION

; APPLICANT: Murphy, Dennis

; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND

; FILE REFERENCE: DIVER1120-4

; CURRENT APPLICATION NUMBER: US/10/112,418

; CURRENT FILING DATE: 2002-03-29

; PRIOR APPLICATION NUMBER: 09/886,400

; PRIOR FILING DATE: 2001-06-20

; PRIOR APPLICATION NUMBER: 09/619,032

; PRIOR FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 09/407,806

; PRIOR FILING DATE: 1999-09-20

; PRIOR APPLICATION NUMBER: 08/613,220

; PRIOR FILING DATE: 1996-03-08

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 364

; TYPE: PRT

; ORGANISM: Thermococcus alcaliphilus

US-10-112-418-4

Query Match

Best Local Similarity 99.4%; Score 1868; DB 9; Length 364;

Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVPHGNLQYAEIPKSEIPVETLKEEIPGLNTGYTLKFLPKDIID 60

Db 1 LRALVPHGNLQYAEIPKSEIPVETLKEEIPGLNTGYTLKFLPKDIID 60

QY 61 LVKGGIADSLIEITIGTSYTHAILPLPLSRVQAQVQDRVKEELFEVSPKGFWLPYLA 120

Db 61 LVKGGIADSLIEITIGTSYTHAILPLPLSRVQAQVQDRVKEELFEVSPKGFWLPYLA 120

QY 121 DPTIIPAILKONGVEYLFADGEAMLFSAHLNSAIKPKIPLYPHLIKQAKREKFRYISYLLG 180

Sequence 51, Appli

Sequence 9, Appli

Sequence 1, Appli

Sequence 5, Appli

Sequence 11997, A

Sequence 21, Appli

Sequence 12884, A

Sequence 5719, Ap

Sequence 12480, A

Sequence 5666, Ap

Sequence 12207, A

Sequence 6, Appli

Sequence 11, Appli

Sequence 12643, A

Sequence 4982, Ap

Sequence 11622, A

Sequence 13, Appli

Sequence 58, Appli

Sequence 59, Appli

Sequence 534, App

Sequence 13801, A

Sequence 11, Appli

Sequence 13418, A

Db 121 DPTIIPAILKONGEYELFADGEAMLFSAHLSAIPKIPKIPYPLHLIKAQREKFRYISYLLG 180
QY 181 LRELKRAIKLVFEGKVTILKAVKDIEAVPVWAVNTAVMLGICRLPLMNPCKVASWIEDKD 240
Db 181 LRELKRAIKLVFEGKVTILKAVKDIEAVPVWAVNTAVMLGICRLPLMNPCKVASWIEDKD 240
QY 241 NILLYGTDIEFTGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db 241 NILLYGTDIEFTGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
QY 301 DKSRLIWRDEGNARLNMLSYNMRGELAFLEAENSADARGWEPLPERRLDRAFRAYINDWRGE 360
Db 301 DKSRLIWRDEGNARLNMLSYNMRGELAFLEAENSADARGWEPLPERRLDRAFRAYINDWRGE 360
QY 361 NGEF 364
Db 361 NGEF 364

RESULT 2

US-10-114-083-4
; Sequence 4, Application US/10114083
; Patent No. US200201604641
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/10/114,083
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus
US-10-114-083-4

Query Match 99.4%; Score 1868; DB 9; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLOYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
Db 1 LRALVFHGNLOYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPYELAY 120
Db 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPYELAY 120
QY 121 DPTIIPAILKONGEYELFADGEAMLFSAHLSAIPKIPKIPYPLHLIKAQREKFRYISYLLG 180
Db 121 DPTIIPAILKONGEYELFADGEAMLFSAHLSAIPKIPKIPYPLHLIKAQREKFRYISYLLG 180
QY 181 LRELKRAIKLVFEGKVTILKAVKDIEAVPVWAVNTAVMLGICRLPLMNPCKVASWIEDKD 240
Db 181 LRELKRAIKLVFEGKVTILKAVKDIEAVPVWAVNTAVMLGICRLPLMNPCKVASWIEDKD 240
QY 241 NILLYGTDIEFTGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db 241 NILLYGTDIEFTGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300

QY 301 DKSRLIWRDEGNARLNMLSYNMRGELAFLEAENSADARGWEPLPERRLDRAFRAYINDWRGE 360
Db 301 DKSRLIWRDEGNARLNMLSYNMRGELAFLEAENSADARGWEPLPERRLDRAFRAYINDWRGE 360
QY 361 NGEF 364
Db 361 NGEF 364

RESULT 3

US-09-886-400-4
; Sequence 4, Application US/09886400
; Patent No. US20020045226A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus
US-09-886-400-4

Query Match 99.4%; Score 1868; DB 10; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLOYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
Db 1 LRALVFHGNLOYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPYELAY 120
Db 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPYELAY 120
QY 121 DPTIIPAILKONGEYELFADGEAMLFSAHLSAIPKIPKIPYPLHLIKAQREKFRYISYLLG 180
Db 121 DPTIIPAILKONGEYELFADGEAMLFSAHLSAIPKIPKIPYPLHLIKAQREKFRYISYLLG 180
QY 181 LRELKRAIKLVFEGKVTILKAVKDIEAVPVWAVNTAVMLGICRLPLMNPCKVASWIEDKD 240
Db 181 LRELKRAIKLVFEGKVTILKAVKDIEAVPVWAVNTAVMLGICRLPLMNPCKVASWIEDKD 240
QY 241 NILLYGTDIEFTGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db 241 NILLYGTDIEFTGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
QY 301 DKSRLIWRDEGNARLNMLSYNMRGELAFLEAENSADARGWEPLPERRLDRAFRAYINDWRGE 360
Db 301 DKSRLIWRDEGNARLNMLSYNMRGELAFLEAENSADARGWEPLPERRLDRAFRAYINDWRGE 360
QY 361 NGEF 364
Db 361 NGEF 364

RESULT 4

US-10-112-357-4
; Sequence 4, Application US/10112357
; Patent No. US20020115099A1
; GENERAL INFORMATION:

```

; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; APPLICANT: Ried, John
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE THEREOF
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/10/112,357
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus
US-10-112-357-4

Query Match          99.4%; Score 1868; DB 12; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKKEEIPFGLNITGYTLKFLPKDIID 60
Db 1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEITGTSYTHAILPLPLSRVEAQVQDREVKKEELFEVSPKGFWLPDELAY 120
Db 61 LVKGGIASDLIEITGTSYTHAILPLPLSRVEAQVQDREVKKEELFEVSPKGFWLPDELAY 120
QY 121 DPIIPAILKONGEYELFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRFVSYLLG 180
Db 121 DPIIPAILKONGEYELFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRFVSYLLG 180
QY 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVWVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
Db 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVWVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
QY 241 NILLYGTDFIEFGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db 241 NILLYGTDFIEFGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
QY 301 DKSRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDFAFRAIYNDRGE 360
Db 301 DKSRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDFAFRAIYNDRGE 360
QY 361 NGEF 364
Db 361 NGEF 364

RESULT 5
US-10-114-403-4
; Sequence 4, Application US/10114403
; Patent No. US20020115100A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; APPLICANT: Ried, John
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE THEREOF
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/10/114,403
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19

```

```

; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus
US-10-114-403-4

Query Match          99.4%; Score 1868; DB 12; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKKEEIPFGLNITGYTLKFLPKDIID 60
Db 1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEITGTSYTHAILPLPLSRVEAQVQDREVKKEELFEVSPKGFWLPDELAY 120
Db 61 LVKGGIASDLIEITGTSYTHAILPLPLSRVEAQVQDREVKKEELFEVSPKGFWLPDELAY 120
QY 121 DPIIPAILKONGEYELFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRFVSYLLG 180
Db 121 DPIIPAILKONGEYELFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRFVSYLLG 180
QY 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVWVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
Db 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVWVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
QY 241 NILLYGTDFIEFGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db 241 NILLYGTDFIEFGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
QY 301 DKSRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDFAFRAIYNDRGE 360
Db 301 DKSRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDFAFRAIYNDRGE 360
QY 361 NGEF 364
Db 361 NGEF 364

RESULT 6
US-10-116-606-4
; Sequence 4, Application US/10116606
; Patent No. US20020119515A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; APPLICANT: Ried, John
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE THEREOF
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/10/116,606
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus
US-10-116-606-4

```

NO

Query Match 99.4%; Score 1868; DB 12; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
DB 1 LRALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEICTSYTHAILPLLSRVEAQVQDRVKEELFEVSPKGFWLPYELAY 120
DB 61 LVKGGIASDLIEICTSYTHAILPLLSRVEAQVQDRVKEELFEVSPKGFWLPYELAY 120
QY 121 DPTIIPAILKONGEYELFADGEMLFSAHLNSAIKPIKPLYPHLIKAKREKFRFYISYLLG 180
DB 121 DPTIIPAILKONGEYELFADGEMLFSAHLNSAIKPIKPLYPHLIKAKREKFRFYISYLLG 180
QY 181 LRELKRAIKLVFECKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKVASWIEDKD 240
DB 181 LRELKRAIKLVFECKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKVASWIEDKD 240
QY 241 NILLYGTDIEFIGYRDIAGRMSVEGLEVIDELNSELCLPSELKHSRGRELYLRTSSWAP 300
DB 241 NILLYGTDIEFIGYRDIAGRMSVEGLEVIDELNSELCLPSELKHSRGRELYLRTSSWAP 300
QY 301 DKSIRIWEDEGNARLNMLSYNMRGELAFIAENSADARGWPELPERRLDAFRAIYNDRWGE 360
DB 301 DKSIRIWEDEGNARLNMLSYNMRGELAFIAENSADARGWPELPERRLDAFRAIYNDRWGE 360
QY 361 NGEF 364
DB 361 NGEF 364

RESULT 7

US-10-112-331-4
; Sequence 4, Application US/10112331
; Patent No. US20020119550A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/10/112,331
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US/09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus

US-10-112-331-4

Query Match 99.4%; Score 1868; DB 12; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
DB 1 LRALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEICTSYTHAILPLLSRVEAQVQDRVKEELFEVSPKGFWLPYELAY 120
DB 61 LVKGGIASDLIEICTSYTHAILPLLSRVEAQVQDRVKEELFEVSPKGFWLPYELAY 120

QY 121 DPTIIPAILKONGEYELFADGEMLFSAHLNSAIKPIKPLYPHLIKAKREKFRFYISYLLG 180
DB 121 DPTIIPAILKONGEYELFADGEMLFSAHLNSAIKPIKPLYPHLIKAKREKFRFYISYLLG 180
QY 181 LRELKRAIKLVFECKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKVASWIEDKD 240
DB 181 LRELKRAIKLVFECKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKVASWIEDKD 240
QY 241 NILLYGTDIEFIGYRDIAGRMSVEGLEVIDELNSELCLPSELKHSRGRELYLRTSSWAP 300
DB 241 NILLYGTDIEFIGYRDIAGRMSVEGLEVIDELNSELCLPSELKHSRGRELYLRTSSWAP 300
QY 301 DKSIRIWEDEGNARLNMLSYNMRGELAFIAENSADARGWPELPERRLDAFRAIYNDRWGE 360
DB 301 DKSIRIWEDEGNARLNMLSYNMRGELAFIAENSADARGWPELPERRLDAFRAIYNDRWGE 360
QY 361 NGEF 364
DB 361 NGEF 364

RESULT 8

US-10-112-377-4
; Sequence 4, Application US/10112377
; Patent No. US20020120108A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/10/112,377
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus

US-10-112-377-4

Query Match 99.4%; Score 1868; DB 12; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
DB 1 LRALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEICTSYTHAILPLLSRVEAQVQDRVKEELFEVSPKGFWLPYELAY 120
DB 61 LVKGGIASDLIEICTSYTHAILPLLSRVEAQVQDRVKEELFEVSPKGFWLPYELAY 120
QY 121 DPTIIPAILKONGEYELFADGEMLFSAHLNSAIKPIKPLYPHLIKAKREKFRFYISYLLG 180
DB 121 DPTIIPAILKONGEYELFADGEMLFSAHLNSAIKPIKPLYPHLIKAKREKFRFYISYLLG 180
QY 181 LRELKRAIKLVFECKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKVASWIEDKD 240
DB 181 LRELKRAIKLVFECKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKVASWIEDKD 240
QY 241 NILLYGTDIEFIGYRDIAGRMSVEGLEVIDELNSELCLPSELKHSRGRELYLRTSSWAP 300
DB 241 NILLYGTDIEFIGYRDIAGRMSVEGLEVIDELNSELCLPSELKHSRGRELYLRTSSWAP 300

QY 301 DKSIRWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
 Db 301 DKSIRWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
 QY 361 NGE 364
 Db 361 NGE 364

RESULT 9
 US-10-116-581-4
 : Sequence 4, Application US/10116581
 : Patent No. US20020137116A1
 : GENERAL INFORMATION:
 : APPLICANT: DIVERSA CORPORATION
 : APPLICANT: Murphy, Dennis
 : APPLICANT: Ried, John
 : TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE THEREOF
 : FILE REFERENCE: DIVER1120-4
 : CURRENT APPLICATION NUMBER: US/10/116,581
 : PRIOR FILING DATE: 2001-04-03
 : PRIOR APPLICATION NUMBER: 09/886,400
 : PRIOR FILING DATE: 2001-06-20
 : PRIOR APPLICATION NUMBER: 09/619,032
 : PRIOR FILING DATE: 2000-07-19
 : PRIOR APPLICATION NUMBER: 09/407,806
 : PRIOR FILING DATE: 1999-09-20
 : PRIOR APPLICATION NUMBER: 08/613,220
 : PRIOR FILING DATE: 1996-03-08
 : NUMBER OF SEQ ID NOS: 4
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ ID NO 4
 : LENGTH: 364
 : TYPE: PRT
 : ORGANISM: Thermococcus alcaliphilus

US-10-116-581-4
 Query Match 99.4%; Score 1868; DB 12; Length 364;
 Best Local Similarity 99.2%; Pred. No. 1.6e-164;
 Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MRALVFHGNLQYAEIPKSEIPKVEKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
 Db 1 LRALVFHGNLQYAEIPKSEIPKVEKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
 QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRVEKVEELFEVSPKGFWLPELAY 120
 Db 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRVEKVEELFEVSPKGFWLPELAY 120
 QY 121 DPTIIPAILKDNQYELFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRFYISYLLG 180
 Db 121 DPTIIPAILKDNQYELFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRFYISYLLG 180
 QY 181 LRELRAIKLVFEGKVTLKAVKDIEAVPVWAVNTAVMLGIGRLPLMNPKKVASWIEDKO 240
 Db 181 LRELRAIKLVFEGKVTLKAVKDIEAVPVWAVNTAVMLGIGRLPLMNPKKVASWIEDKO 240
 QY 241 NILLYGTDFIEFGYRDYAGYRMSVEGLLEVIDELNSELCLPSKLGHSRELYLRTSSWAP 300
 Db 241 NILLYGTDFIEFGYRDYAGYRMSVEGLLEVIDELNSELCLPSKLGHSRELYLRTSSWAP 300
 QY 301 DKSIRWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
 Db 301 DKSIRWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
 QY 361 NGE 364
 Db 361 NGE 364

RESULT 10
 US-10-112-442-4

: Sequence 4, Application US/10112442
 : Patent No. US20020150997A1
 : GENERAL INFORMATION:
 : APPLICANT: DIVERSA CORPORATION
 : APPLICANT: Murphy, Dennis
 : APPLICANT: Ried, John
 : TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE THEREOF
 : FILE REFERENCE: DIVER1120-4
 : CURRENT APPLICATION NUMBER: US/10/112,442
 : PRIOR FILING DATE: 2002-03-29
 : PRIOR APPLICATION NUMBER: 09/886,400
 : PRIOR FILING DATE: 2001-06-20
 : PRIOR APPLICATION NUMBER: 09/619,032
 : PRIOR FILING DATE: 2000-07-19
 : PRIOR APPLICATION NUMBER: 09/407,806
 : PRIOR FILING DATE: 1999-09-20
 : PRIOR APPLICATION NUMBER: 08/613,220
 : PRIOR FILING DATE: 1996-03-08
 : NUMBER OF SEQ ID NOS: 4
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ ID NO 4
 : LENGTH: 364
 : TYPE: PRT
 : ORGANISM: Thermococcus alcaliphilus

US-10-112-442-4
 Query Match 99.4%; Score 1868; DB 12; Length 364;
 Best Local Similarity 99.2%; Pred. No. 1.6e-164;
 Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MRALVFHGNLQYAEIPKSEIPKVEKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
 Db 1 LRALVFHGNLQYAEIPKSEIPKVEKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
 QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRVEKVEELFEVSPKGFWLPELAY 120
 Db 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRVEKVEELFEVSPKGFWLPELAY 120
 QY 121 DPTIIPAILKDNQYELFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRFYISYLLG 180
 Db 121 DPTIIPAILKDNQYELFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRFYISYLLG 180
 QY 181 LRELRAIKLVFEGKVTLKAVKDIEAVPVWAVNTAVMLGIGRLPLMNPKKVASWIEDKO 240
 Db 181 LRELRAIKLVFEGKVTLKAVKDIEAVPVWAVNTAVMLGIGRLPLMNPKKVASWIEDKO 240
 QY 241 NILLYGTDFIEFGYRDYAGYRMSVEGLLEVIDELNSELCLPSKLGHSRELYLRTSSWAP 300
 Db 241 NILLYGTDFIEFGYRDYAGYRMSVEGLLEVIDELNSELCLPSKLGHSRELYLRTSSWAP 300
 QY 301 DKSIRWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
 Db 301 DKSIRWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
 QY 361 NGE 364
 Db 361 NGE 364

RESULT 11
 US-09-815-242-12361
 : Sequence 12361, Application US/09815242
 : Patent No. US20020061569A1
 : GENERAL INFORMATION:
 : APPLICANT: Haselbeck, Robert
 : APPLICANT: Ohlsen, Kari L.
 : APPLICANT: Zyskind, Judith W.
 : APPLICANT: Wall, Daniel
 : APPLICANT: Carr, Grant J.
 : APPLICANT: Yamamoto, Robert T.
 : APPLICANT: Xu, H. Howard
 : TITLE OF INVENTION: Identification of Essential Genes in

```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12361
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12361

```

```

Query Match          5.5%; Score 103; DB 10; Length 1073;
Best Local Similarity 21.2%; Pred. No. 0.28;
Matches 58; Conservative 48; Mismatches 90; Indels 78; Gaps 15;

QY 107 EVSPKGFWLPDELAYDPIIPAILKONGYEYLFADGEAMLSAHLNSAIKPKPLYPHLIKA 166
Db 309 EISP-----YDLSL---LVKLSHAISFKQAEKVMRSLREMRIRGVKTNIPFLINV 357
QY 167 QREKRF-----RYISYLLGLRLRKAI-----KLVFEGKVTILKAVKDIE----- 205
Db 358 MNKFKFTSGDYTTKFTETPELFDIQPSLDGRTKLTLEYIGNVTINGFPNVEKRPKPDYEL 417
QY 206 -AVPWWAVNTAVMLGIGR-LPLMNPKKVASWIEDKDNILLYGTDIEF-IGYRDIAGYRM 262
Db 418 ASIPTVSSSKIASFSGKQLLDEVGPKGVAEWKKQDDVLL--TDTTFRDAHQSLLATRV 475
QY 263 SVEGLEVIDELNSELCLPSKLSKHSRELYLRTSSWAPDK-SLRWREDEGNARLNLSY 321
Db 476 RTKDMINIAS-----KTADVFKDGFSLMWM-----GGATFD-VAY 509
QY 322 NMRGELAFLAENSARGWEPLPERRLDARAIYN 355
Db 510 N-----FLKENP-----WERLERLR-----KAIPN 529

```

```

RESULT 12
US-09-815-242-5468
; Sequence 5468, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

```

```

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5468
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5468

```

```

Query Match          5.5%; Score 103; DB 10; Length 1147;
Best Local Similarity 21.2%; Pred. No. 0.31;
Matches 58; Conservative 48; Mismatches 90; Indels 78; Gaps 15;

QY 107 EVSPKGFWLPDELAYDPIIPAILKONGYEYLFADGEAMLSAHLNSAIKPKPLYPHLIKA 166
Db 385 EISP-----YDLSL---LVKLSHAISFKQAEKVMRSLREMRIRGVKTNIPFLINV 433
QY 167 QREKRF-----RYISYLLGLRLRKAI-----KLVFEGKVTILKAVKDIE----- 205
Db 434 MNKFKFTSGDYTTKFTETPELFDIQPSLDGRTKLTLEYIGNVTINGFPNVEKRPKPDYEL 493
QY 206 -AVPWWAVNTAVMLGIGR-LPLMNPKKVASWIEDKDNILLYGTDIEF-IGYRDIAGYRM 262
Db 494 ASIPTVSSSKIASFSGKQLLDEVGPKGVAEWKKQDDVLL--TDTTFRDAHQSLLATRV 551
QY 263 SVEGLEVIDELNSELCLPSKLSKHSRELYLRTSSWAPDK-SLRWREDEGNARLNLSY 321
Db 552 RTKDMINIAS-----KTADVFKDGFSLMWM-----GGATFD-VAY 585
QY 322 NMRGELAFLAENSARGWEPLPERRLDARAIYN 355
Db 586 N-----FLKENP-----WERLERLR-----KAIPN 605

```

```

RESULT 13
US-09-732-180-7
; Sequence 7, Application US/09732180
; Patent No. US20020137702A1
; GENERAL INFORMATION:
; APPLICANT: Arkowitz, Robert A
; APPLICANT: Nern, Peter MA
; TITLE OF INVENTION: Yeast receptor
; FILE REFERENCE: DYOU25.001AUS
; CURRENT APPLICATION NUMBER: US/09/732,180
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,699
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-732-180-7

```

```

Query Match          4.8%; Score 91; DB 10; Length 813;
Best Local Similarity 22.1%; Pred. No. 2.4;
Matches 81; Conservative 56; Mismatches 133; Indels 96; Gaps 19;

QY 25 EKAYIPVIETLIK--EETPFGINITGYTKFL---PKDIIDLVK---GGIASDL-----I 71
Db 275 ERKYVQDLELCKYKRDLEAENLSSEQIHLLFNLNFIIDFORFLNGLNENVPYRY 334
QY 72 ELIGTSYTHAIL-----PFLPLSRVEAQVORDREVKEELFEVSPKGFWLPDELAYDP 122

```


Db 335 QRGSVFIHSLGPNAYEPWTIGQLTAIDLINKEAANKKSSSLDDP-GFELQSYILKP 393
Qy 123 I-----IPAILKD---NGEYLFADGAMLFSAHLNS---AIKPIKPLPHLIKAQREKR 171
Db 394 IQRLCKYPLLLKELIKTSPEYSKODPHGSSSTSFNELLVAKTAMKELANOVNEAQR--R 451
Qy 172 FRYISYLLGLRE-----LRKAIKLVFEGKVTLKAVKDIEAVPWWAVNTAVMLGIG 222
Db 452 AENIEHLEKUKERVGNWRGNLDAQGELLPHGVQV---GVKDAE----- 491
Qy 223 RLPLMNPKKVASWIEDKDNILLYGTDIEFTGYRD-----IAGYRMSVEGL 267
Db 492 -----NEKEYVAYLFER--IVFFTEIDDTKKSDQEKSKFSTRKRSTSSNLSSSTTNL 544
Qy 268 LEVIDELNSELCLPSELKHSR-----ELYLRTSSWAPDKSLRI-W--REDEGNARLNMLS 320
Db 545 LESINNSRKONTLPLELK--GRVYISYINISAPNTPGSLTIISWSGRKESGFTLYRS 602
Qy 321 YNMGRGE 326
Db 603 EARNQ 608
RESULT 14
US-09-732-180-2
; Sequence 2, Application US/09732180
; Patent No. US20020137702A1
; GENERAL INFORMATION:
; APPLICANT: Arkowitz, Robert A
; APPLICANT: Nern, Peter MA
; TITLE OF INVENTION: Yeast receptor
; FILE REFERENCE: DYO025.001AUS
; CURRENT APPLICATION NUMBER: US/09/732,180
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,699
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 844
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-732-180-2

Query Match 4.8%; Score 91; DB 10; Length 844;
Best Local Similarity 22.1%; Pred. No. 2.6;
Matches 81; Conservative 56; Mismatches 133; Indels 96; Gaps 19;
Qy 25 EKAVIPVETLIK--EEIPIFGLNITGYTLKFL---PKDIIDLVK---GGIASDL-----I 71
Db 289 ERKVQDLELMCKYRQDLIEAENLSSQIHLLFPNLNEIDFORFLNGLECNINVPYR 348
Qy 72 EICTSTYTHAIL-----PPLPLSRVAQVQREVKEELFVSPKGFWLPFLAYDP 122
Db 349 QRGSVFIHSLGPNAYEPWTIGQLTAIDLINKEAANKKSSSLDDP-GFELQSYILKP 407
Qy 123 I-----IPAILKD---NGEYLFADGAMLFSAHLNS---AIKPIKPLPHLIKAQREKR 171
Db 408 IQRLCKYPLLLKELIKTSPEYSKODPHGSSSTSFNELLVAKTAMKELANOVNEAQR--R 465
Qy 172 FRYISYLLGLRE-----LRKAIKLVFEGKVTLKAVKDIEAVPWWAVNTAVMLGIG 222
Db 466 AENIEHLEKUKERVGNWRGNLDAQGELLPHGVQV---GVKDAE----- 505
Qy 223 RLPLMNPKKVASWIEDKDNILLYGTDIEFTGYRD-----IAGYRMSVEGL 267
Db 506 -----NEKEYVAYLFER--IVFFTEIDDTKKSDQEKSKFSTRKRSTSSNLSSSTTNL 558
Qy 268 LEVIDELNSELCLPSELKHSR-----ELYLRTSSWAPDKSLRI-W--REDEGNARLNMLS 320
Db 559 LESINNSRKONTLPLELK--GRVYISYINISAPNTPGSLTIISWSGRKESGFTLYRS 616

Qy 321 YNMGRGE 326
Db 617 EARNQ 622
RESULT 15
US-09-815-242-5654
; Sequence 5654, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5654
; LENGTH: 1668
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5654

Query Match 4.8%; Score 89.5; DB 10; Length 1668;
Best Local Similarity 19.6%; Pred. No. 9.2;
Matches 67; Conservative 68; Mismatches 128; Indels 79; Gaps 15;
Qy 17 KSEIPKVIKAVIPVETLIKKEIIPFGLN-----ITGYTLK 52
Db 513 KSQRLMTLPEYMLPV-NFMHIEQIPITINGKLDKALPIMDYDVTDAVAPSTDTEHLIC 571
Qy 53 FLPKDIDLVKGGIASDLIEIIGTSYTHAILPPLSRVAQVQREV-----KEELFE 107
Db 572 QIFADILLHVNOVGIIHDFNFFELGG---HSLKATLVVNRIEASTGRKQLQIGLLQKPTVPE 627
Qy 108 VSPKGFWLPFLAYDPIIPAILKDNQYVEYLFADGEAMLF-----SAHLNSAIKPIKPLPHL 163
Db 628 LAQATIAKVOEQNYEYIPEAIVKD---DYVLSSAQKRMVLLWKSNNHKTYYN-----VPFL 679
Qy 164 IKAQREKRE---RVISYLLGLRELKSKALKVFEKVTLLKAVKDIEAVPWWAVNT----- 215
Db 680 WRLSSSELNVAQLRQAVQHLIARHEILRTQYIVVDDEVQRQIVADV--VADFEVWTHFTD 737
Qy 216 ---AVMLGIGRLPLMNPKKV-ASWIEDKDNILLYGTDIEFTGYRDIAGYRMSVEGLEVI 271
Db 738 EQEIMRQFVAPFNLEKPSQIRVRYIRSPHLAVL-----FIDTHHINDGMS---NIQLM 788
Qy 272 DELNS-----ELCLPSELKHSRELYLRTSSWAPDKSLRIWRE 309
Db 789 NDLNALLYQHLLPLKLVQ-----YKDYSEWMSHRDMTKHRQ 824

Mon Nov 18 08:39:35 2002

us-09-619-032-4.rapb

Page 8

Search completed: November 18, 2002, 07:30:03
Job time : 11 secs